R E P O R T R E S U M E S

ED 017 485 TE 000 223

AN EVALUATION OF PUBLISHED ENGLISH TESTS.

BY- WOOD, SUSAN

WISCONSIN STATE DEPT. OF PUB. INSTR., MADISON
REPORT NUMBER DPI-BULL-144 PUB DATE 67
EDRS PRICE MF-$0.50 HC-$3.44 84P.

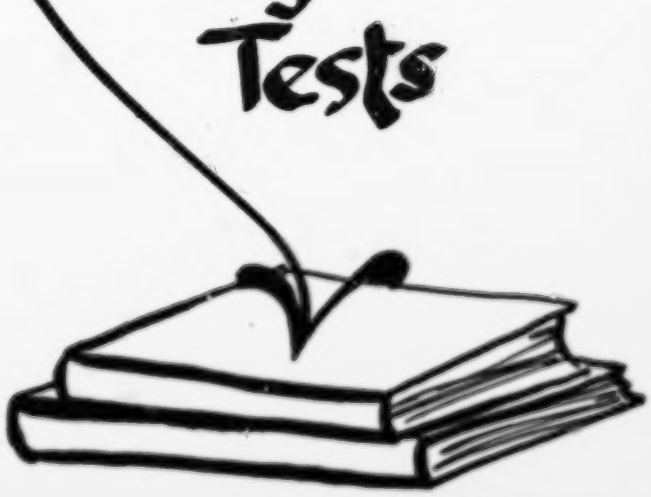
DESCRIPTORS- *ELEMENTARY SCHOOLS, *ENGLISH INSTRUCTION, *LANGUAGE TESTS, *SECONDARY SCHOOLS, *STANDARDIZED TESTS, TESTS, ACHIEVEMENT TESTS, ESSAY TESTS, GROUP TESTS, APTITUDE TESTS, OBJECTIVE TESTS, VERBAL TESTS, LANGUAGE PROFICIENCY,

THIS STUDY, CONDUCTED BY THE WISCONSIN ENGLISH-LANGUAGE-ARTS CURRICULUM PROJECT, DESCRIBES, ANALYZES, AND EVALUATES 16 STANDARDIZED ENGLISH TESTS OF USAGE AND COMPOSITION. THE TESTS CHOSEN WERE THOSE FREQUENTLY USED THROUGHOUT WISCONSIN, EXCLUSIVE OF TESTS DESIGNED FOR COLLEGE-PREPARATORY STUDENTS AND TESTS OF READING ABILITY, SPEECH, AND LITERATURE. ALTHOUGH MANY OF THE TESTS EVALUATED ARE RESTRICTED TO ENGLISH SKILLS, SOME ARE LANGUAGE SUB-TESTS OF MULTI-SUBJECT ACHIEVEMENT BATTERIES. THOSE EVALUATED ARE--(1) THE BARRETT-RYAN-SCHRAMMEL ENGLISH TEST, NEW EDITION, (2) CALIFORNIA LANGUAGE TEST, (3) COOPERATIVE ENGLISH TESTS, 1960 REVISION, (4) DIFFERENTIAL APTITUDE TESTS, (5) ESSENTIALS OF ENGLISH TESTS, REVISED EDITION, (6) GREENE-STAPP LANGUAGE ABILITIES TEST, (7) IOWA TESTS OF BASIC SKILLS, (8) IOWA TESTS OF EDUCATIONAL DEVELOPMENT, (9) METROPOLITAN ACHIEVEMENT TESTS, (10) OBJECTIVE TEST IN GRAMMAR, (11) PURDUE HIGH SCHOOL ENGLISH TEST, (12) COOPERATIVE SCHOOL AND COLLEGE ABILITY TESTS, (13) SCIENCE RESEARCH ASSOCIATES (SRA) ACHIEVEMENT SERIES--LANGUAGE ARTS, (14) SRA HIGH SCHOOL PLACEMENT TEST, (15) SEQUENTIAL TESTS OF EDUCATIONAL PROGRESS, (16) AND STANFORD ACHIEVEMENT TEST, 1964 REVISION. EACH TEST IS DISCUSSED UNDER FOUR HEADS--GENERAL INFORMATION, USE IN WISCONSIN, TEACHER EVALUATIONS, AND PUBLISHED REVIEWS. ALSO INCLUDED ARE SOME CONCLUSIONS ABOUT THE ADEQUACY OF STANDARDIZED ENGLISH TESTS IN GENERAL AND THE PROCESS OF TEST SELECTION, A LIST OF SIX TESTS FOUND USEFUL IN WISCONSIN, AND A LIST OF TEST PUBLISHERS. THIS BULLETIN IS AVAILABLE FROM THE PUBLICATIONS ORDER DIVISION, DEPARTMENT OF PUBLIC INSTRUCTION, 126 LANGDON STREET, MADISON, WISCONSIN 53702, $0.75. (MM)

ED017485

TE 000 223

Evaluation of Published English Tests



WISCONSIN DEPARTMENT OF PUBLIC INSTRUCTION
MADISON, WISCONSIN 53702

William C. Kahl, State Superintendent

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

An Evaluation of Published English Tests

Susan Wood
under the direction of
Dr. Robert C. Pooley

A guide for administrators, supervisors and teachers of English in the selection and use of standardized tests. Sixteen tests used in the elementary and secondary schools of Wisconsin are described, analyzed and evaluated.

DPI Bulletin No. 144
Wisconsin English-Language-Arts
Curriculum Project
5000—21

Contents

Page

Introduction 3

Each test is discussed under four heads:
- I. General Information
- II. Use in Wisconsin
- III. Teacher Evaluations
- IV. Published Reviews

Barrett-Ryan-Schrammel English Test, New Edition 9

California Language Test 13

Cooperative English Tests, 1960 Revision 19

Differential Aptitude Tests 23

Essentials of English Tests, Revised Edition 27

Greene-Stapp Language Abilities Test 31

Iowa Tests of Basic Skills 35

Iowa Tests of Educational Development 41

Metropolitan Achievement Tests 47

Objective Test in Grammar 55

The Purdue High School English Test 57

(Cooperative) School and College Ability Tests 61

SRA Achievement Series: Language Arts 65

SRA High School Placement Test 69

Sequential Tests of Educational Progress 73

Stanford Achievement Test [1964 Revision] 79

Conclusions 85

Recommendations 87

Test Publishers 88

Introduction

In recent years Wisconsin educators have voiced concern about the adequacy of commercially available standardized English tests. In addition, many teachers and administrators have expressed a desire to learn more about such tests. What tests are actually available and what do they measure? How do they compare with each other? And most important, how can they be useful to the classroom teacher? In this study the Wisconsin English-Language-Arts Curriculum Project hopes to answer these questions and to raise several others. For instance, can an objective test effectively measure writing ability? To what standards of usage should such tests subscribe? Controversial questions such as these can never be answered to everyone's satisfaction. It is hoped, however, that the debate they promote will define the various positions being taken on these important issues. If the survey succeeds in performing this informative service to Wisconsin educators, it will have fulfilled its purpose.

Work on this study began in November, 1965. Immediately it became apparent that discussion of all standardized English tests in a bulletin of limited scope was a physical impossibility. Therefore, the subject was narrowed in three ways. First of all, tests of reading ability, speech, and literature, which seem to have aroused little controversy among educators, were excluded. It was decided that tests of usage and composition, which have been the subject of much debate, offered a more suitable field of investigation. Under these general headings were included all measurements of spelling, vocabulary, sentence structure, grammatical awareness, and mechanics, as well as specific tests of word usage and writing ability. Language subtests included in multi-subject achievement batteries, many of which are available separately, were considered in addition to tests confined exclusively to English skills. Secondly, tests designed to be administered only to college preparatory students (e.g., the *National Merit Scholarship Qualifying Test*, the *American College test*, etc.) were excluded. It was felt that although such instruments do measure usage and composition, their specific purpose disqualifies them from a survey of this nature. Tests measuring the speaking and writing needs and abilities of *all* students are the proper subject of such a study. Finally, since the

survey was intended to assist Wisconsin educators, it was decided that only those usage and composition tests which are in actual and frequent use throughout the state would be considered.

In order to discover which tests are used in Wisconsin, a questionnaire was mailed to the 425 district administrators in February, 1966. In March, duplicates of this questionnaire were sent to those administrators who had not yet responded. Hereafter this questionnaire will be referred to as the first questionnaire. Three hundred and thirty-one, or 78 per cent of the administrators returned the questionnaire. Fourteen responses were multiple: i.e., the administrator duplicated the questionnaire and distributed it to several teachers for completion. Seventy-two additional responses were attained in this manner, raising the total number of respondents to 403. Ninety-seven, or 21 per cent of these respondents do not employ standardized tests of usage or composition (these figures include those who use only standardized reading, speech, or literature tests and those who employ only tests accompanying textbooks). Thus, 306, or 79 per cent of the respondents use standardized usage and composition tests. Seventy-six, or 19 percent, use tests which accompany textbooks, either exclusively or to supplement the results of standardized tests.

It has been noted that many administrators submitted the first questionnaire to an associate or associates for completion. The following table presents the positions of those who completed the questionnaire in terms of numerical value and percentage of the total:

Table 1

School Position	Number	Percentage of Total
School Psychologists	3	1%
Curriculum Supervisors and Coordinators	29	7%
School Superintendents	42	10%
Guidance Counselors	47	12%
Principals	56	14%
English Teachers	193	48%
Administrators	29	7%
Undesignated	4	1%
Total	403	100%

First Questionnaire

The first questionnaire contained the following questions:

1. Which objective tests of English skills have you used in in the last five years?
2. Of these, which have you found satisfactory and (in a word or two)why?
3. Which have proved unsatisfactory and (in a word or two) why?
4. Do objective tests measure what your teachers consider important in the language arts?

Respondents were also invited to include other questions and comments.

The following table lists those standardized tests of usage and composition which are in widest use in Wisconsin according to the results of the first questionnaire. Figures are presented in numerical values and in terms of approximate percentages of the total number of respondents (306) who use such tests.

Table 2

Test	Number Using It	Percentage Using It	Abbreviation
Barrett-Ryan-Schrammel English Test	4	1%	BRS
California Language Test	25	8%	CLT
Cooperative English Tests	11	4%	CET
Differential Aptitude Tests	6	2%	DAT
Essentials of English Tests	16	5%	EE
Greene-Stapp Language Abilities Test	8	3%	GS
Iowa Tests of Basic Skills	123	40%	ITBS
Iowa Tests of Educational Development	61	20%	ITED
Metropolitan Achievement Tests	45	15%	MAT
Objective Test in Grammar	5	2%	OTG
Purdue High School English Test	8	3%	PHET
School & College Ability Tests	24	8%	SCAT
Science Research Associates: Language Arts	15	5%	SRA
Science Research Associates: High School Placement Test	7	2%	SRA-HPT
Sequential Tests of Educational Progress	103	34%	STEP
Stanford Achievement Tests	32	10%	SAT

Of course, many other standardized tests of usage and composition are employed in Wisconsin. Those listed above appear to be the most popular and will be discussed in this survey. Although they are probably in wider use than the figures indicate, it is reasonable to assume that Table 2 conveys a fairly accurate picture of the general trends of usage and popularity. (The third column presents alphabetical abbreviations which will be substituted throughout this report for the full title of each test.)

The results of Questions 2 and 3 of the first questionnaire will be included in the discussions of individual tests. Question 4, "Do objective tests measure what your teachers consider important in the language arts?" was answered in the following fashion:

Table 3

Response	Number
Yes (meaning objective tests in general)	47
Yes (meaning those objective tests actually used)	55
Qualified Yes	45
No (meaning objective tests in general)	72
No (meaning those objective tests actually used)	28
Qualified No	24
No Answer	132

Six respondents considered Question 4 difficult to answer because teachers disagree as to what is "important" in the language arts. Six others indicated that English teachers are unaware of or uninterested in the results of standardized tests, either because they are badly informed or because they place little value on the results. Fourteen considered the coverage of grammar on such tests inadequate or outdated, and interestingly, 41 of those who answered "no" did so because they considered objective tests an inadequate measure of writing ability and creativity. One respondent expressed his own doubts and those of many others in this comment: "We are realizing more and more that correctness (absence of actual errors) is not enough. I am wondering now if isolated sentences can test for excellence in the use of the English language." Another's objection was also frequently repeated: "Why do tests continue to quiz students on details of traditional grammar, despite recent criticisms of teaching it?"

Under Question 5, which invited additional comments, many respondents expressed awareness of the need for sweeping improve-

ment of standardized English tests. They noted that the inadequacy of many tests not only hampers assessment of student needs and abilities, but adversely affects the curriculums of schools that "teach to the tests." Several respondents considered the tests useful only for the general purposes of placement, general guidance, and comparison of schools. Others judged test results useful if supplemented by the results of teacher-made tests, classroom work, and original writing. Several interesting points were made by individual respondents. One noted that since high scores on many tests are directly related to speed of performance, the thorough but slow reader is often penalized. Another commented that persistently low scores on English tests "cast their reflections against the department and add considerably to the trials of teaching." Are teachers really trying to produce the kind of learning that present tests are designed to measure? Do the majority of standardized English tests test only memorization of rules and isolated facts? If reliance upon rules is undesirable, what is the answer? A third respondent suggested that tests be designed to measure the acquisition of concepts relevant to the "new" grammar and its application. On the other hand, another respondent defined as the proper role of standardized tests the measurement of purely mechanical skills. Finally, one teacher suggested a one-day meeting of state administrators and curriculum coordinators to explore the merits of standardized testing.

These, then, were the results of the first questionnaire. During the spring of 1966, specimen sets of each test listed in Table 2 were ordered and received from their respective publishers. Next, the aid of elementary and secondary English teachers representing schools throughout Wisconsin was enlisted. (See acknowledgements pp. 90-91.) Each teacher agreed to evaluate one test designed for his grade level and received a test specimen set and a second questionnaire to complete. Requests to evaluate specific tests were approved whenever possible. The following questions were included on this questionnaire:

Second Questionnaire

1. From your point of view as a teacher, what are the chief strengths and weaknesses of this test?
2. In your estimation, is the test successful in measuring the abilities it claims to measure?

3. Does the test content include material which you consider a valid part of the English curriculum?
4. Is there a direct relation between the content of the test and the dialect of "informal standard English" which you are trying to establish in your classroom?
5. Are you confident that the test is long and comprehensive enough to provide a reliable indication of student aptitude and/or achievement?
6. Do you feel that the stated norms for this test would provide a realistic guide for measuring the performance of *your* students?
7. Are the supplementary test materials (answer sheets, directions for administering, guide and key to scoring, etc.) clear and comprehensive?
8. Would you actually use this test in your classroom?
9. Do you use it now?

The teachers were instructed to take the tests themselves before completing the questionnaires. Responses to these questions are presented as part of the discussions of individual tests.

Finally, published reviews of the most recent edition of each test were consulted. (All reviews summarized in this survey may be found in Oscar K. Buros's *The Fifth Mental Measurements Yearbook*, Highland Park, N. J.: The Gryphon Press, 1959, or in *The Sixth Mental Measurements Yearbook*, Highland Park, N. J.: The Gryphon Press, 1965.) Unless otherwise indicated, references in this study are to *The Sixth Mental Measurements Yearbook*.

This survey presents the following information about each test:

1. Bibliographical information and a general description.
2. A summary of the test's use in Wisconsin, as indicated by the first questionnaire.
3. An evaluation by one or more Wisconsin English teachers.
4. A synopsis of one or more published reviews.

Barrett-Ryan-Schrammel English Test, New Edition

I. General Information

Grades 9-13; 1938-54; 6 scores: grammar, sentence, punctuation, vocabulary, pronunciation, total; Forms DM, EM ('54), and FM; manual ('54); $3.50 per 35 tests; $1.70 per 35 IBM answer sheets; 60 (70) minutes; E. R. Barrett, Teresa M. Ryan, and H. E. Schrammel; Harcourt, Brace & World, Inc.

The BRS purports to measure objectively proficiency in handling the essentials of English mechanics. The publishers recommend that it be used for diagnostic and survey purposes and for placement in high school and college classes. Each of its three forms contains 179 items based on the common content of leading textbooks and courses of study. The vocabulary and pronunciation tests are new to this edition; other subtests have been revised. The test was standardized by administration to 32,641 high school students and to 7,212 college freshmen representing nationwide distribution. Means and standard deviations of scores are given by test and grade; split-half and alternate form reliability coefficients are provided. Standard errors of measurement and tables of percentile ranks corresponding to part and total scores are also included.

II. Use in Wisconsin

The BRS is used by four, or 1 percent of those respondents to the first questionnaire who employ standard tests. It was rated satisfactory by half of its users, and was judged unsatisfactory by the other two. Those ruling it satisfactory liked its emphasis on "functional grammar," but those rating it unsatisfactory felt it overstressed the nomenclature of formal grammar.

III. Teacher Evaluations

The evaluators of the BRS, Form EM, agreed substantially in their opinions. On the whole, they felt that the test's weaknesses outnumbered its strengths. In the first place, the answer sheet located on the final page of the test proved difficult to locate and

to use. Some items in Part I, furthermore, depend upon the immediately preceding items so that if the examinee is unable to answer one item, he will be unable to answer the succeeding ones. The following discrepancies were cited: use of the term "predicate verb" in Part II (The Sentence), failure to differentiate between transitive and linking verbs, and imprecision in the use of terminology, e.g., "direct object of verb" for "object of infinitive." It was felt also that the omission of sample answers in the vocabulary section and the arbitrariness of determining what may be considered effective choices of sentence constructions constituted major flaws in content.

The evaluators agreed that the test is an adequate index of the skills it claims to measure, but questioned the importance of those skills. They observed that the test does not give enough attention to the skills required in the actual writing process: word choice, sentence openers and transitional phrases, manipulation of basic sentence patterns, etc.

These teachers agreed further that only a portion of the test content is a valid part of the English curriculum. One-half of the group considered the punctuation section is superior, but rated the vocabulary section inadequate for current needs. In addition, this same group objected to the number of items devoted to syllables and accents. The other half felt that the vocabulary and pronunciation sections are adequate as they stand, but thought that misplaced modifiers and lack of parallelism should be tested. These teachers, therefore, located content inadequacy in different places, but definitely agreed that it exists.

The teachers also differed in their estimates of the relation of test content to informal standard English. They agreed, however, that the test is neither long nor comprehensive enough to provide a reliable measure of English achievement. One remarked that the 60 minute working time makes it almost impossible to use the test during a regular classroom period for diagnostic purposes.

One half of these evaluators felt that the stated norms would not provide an adequate guide for measuring the performance of their students; the other half deemed it necessary to compare the norms with actual student scores before commenting upon them. All of these teachers considered the supplementary materials generally clear and comprehensive, although about half of them questioned the clarity of the directions for obtaining an examinee's

total score on the test. None of these teachers had used the test before. About half indicated they would like to try it in the classroom, but the other half's strong objections would prevent them from adopting it.

IV. Published Reviews

In *The Fifth Mental Measurements Yearbook,* Oscar K. Buros, ed., may be found the following reviews:

Leonard S. Feldt, Assistant Professor of Education at the State University of Iowa, gives the BRS an unfavorable review. He can find no evidence that this instrument is a good test of proficiency in English mechanics, and he sees no factual evidence, for example, that "would allow the potential user to evaluate the appropriateness of content" (p. 330). There are insufficient data to back up the authors' claims regarding the use of scores for placement and diagnosis. Furthermore, the content does not include enough items on each problem situation to make the diagnosis of student strengths and weaknesses an accurate one. In Mr. Feldt's opinion, the test includes "far too many items (67) on the academic aspects of language . . . and too few items (52) involving functional mechanics" (p. 331). For instance, there are no items on capitalization and spelling.

Cleveland A. Thomas, Principal of the Francis W. Parker School in Chicago, Illinois, agrees with Mr. Feldt that Part I of the BRS (Functional Grammar) actually tests formal grammar, or "knowledge of grammar in a vacuum, in a way not necessarily related to speech and writing" (p. 331). Like the teachers who evaluated the test, Mr. Thomas accepts the vocabulary and punctuation sections of the test, but criticizes the lack of description of the basis for the selection of vocabulary words. In addition, he feels that the subtest on the sentence "is actually a test of the grammar of the sentence, . . . [not of] the students' skill in the construction of sentences" (p. 331). He suggests that the test might be improved by the inclusion of items in appropriateness and sentence structure similar to those in the *CEEB English Achievement Test,* and by the adoption of fuller contexts for many items. He believes that although the BRS will be of more interest to teachers who teach formal grammar than to those who are concerned with students' speaking and writing ability, even the former group will find its

value limited by the unreliability of part scores. Despite these objections, however, Mr. Thomas feels that the BRS is "as good an overall measure of the mechanics of English as other tests of the same" (p. 332).

California Language Test

I. General Information

1957 Edition with 1963 norms. Grades 1-2, 2-4, 5, 4-6, 7-9, 9-14; 1933-63; subtest of *California Achievement Tests;* 4 scores: mechanics of English, spelling, total, handwriting; IBM and Grade-O-Mat for grades 4-14; 2-4 forms ('63 printings identical with '57 except for profile); manual for each of 5 levels; technical report for '57 edition with '57 norms; individual profile for each level; separate answer sheets available for grades 4-14.

a) Lower Primary (1-2): $2.45 per 35 tests; 27 (40) minutes; Forms W ('63), X ('57).
b) Upper Primary (2.5-4.5): $2.80 per 35 tests; 30 (40) minutes; Forms W ('63), X ('57).
c) Elementary (4-6)*: $3.15 per 35 tests; 40 (50) minutes; Forms W ('63), X ('57), Y ('63).
d) Junior High (7-9)*: $3.15 per 35 tests; 32 (40) minutes; Forms W ('63), X ('57), & Y ('57).
e) Advanced (9-14): $3.15 per 35 tests; 38 (48) minutes; Forms W ('63), X ('57), Y ('57).

Ernest W. Tiegs and Willis W. Clark, California Test Bureau.

According to the manual, the CLT is designed for the measurement, evaluation, and diagnosis of school achievement. The five levels of the battery are designed to provide a sequential testing program from one level to the next. The sequential nature of the test has been retained in the preparation of norms, which are based "on a scaling procedure by which performance on the CAT was related to performance on the 1963 Revision of the *California Short-Form Test of Mental Maturity* . . . ." This test's population sample represents a nation-wide cross-section of curricular trends. Correct answer positions are the same on all forms at the same level, so that one set of keys will score and analyze any form at a given level; hence, only one set of normative data is needed at a given level.

*The manual advises that norms for the total language test and the total battery have been modified as of June, 1965, for grades 4-6; percentile ranks, standard scores, stanines were modified in May, 1964, for grades 7-9.

II. Use in Wisconsin

The CLT is used by 25, or 8 percent, of the standard test users who responded to the first questionnaire. Seven, or 28 percent of the total number of users rated it satisfactory because of its curricular validity and its general outline of proficiencies and deficiencies for use in the development of individualized instruction programs. Nine, or 36 percent, rated it unsatisfactory for various and often conflicting reasons: three considered it too simple; two thought it too difficult; two others suggested it could be more comprehensive; another objected to everything from means of marking and level of difficulty, to norms. Nine users, or 36 percent, did not rate it at all, thereby rendering the above ratings somewhat less than valid.

III. Teacher Evaluations

The CLT, like many other tests, contains separate batteries for each of several grade levels, each battery being treated as a separate entity. Each teacher was asked to evaluate one battery and was later assigned one suitable to the grade level which he teaches. In all, six teachers contributed to the following evaluation (two teachers evaluated the Elementary Battery). For the sake of convenience, the batteries for each of the five levels will be discussed separately. The same procedure will be followed in other sections of this study devoted to similarly constructed tests.

Lower Primary Battery (Form W, Grades 1-2)

No report.

Upper Primary Battery (For W, Grades 2-4)

The teacher evaluating this section of the CLT concluded that it needs much improvement. She suggested inclusion of material covering letter writing, paragraphs, and topic sentences, perhaps in the punctuation and capitalization sections. She considered the content a valid part of the curriculum, but judged the test not comprehensive enough to provide an accurate measure of pupils' true ability. In fact, she suspected that the skills tested are too elementary for her pupils. She also objected to the ambiguity inherent in the nature and form of the supplementary materials, and stated that she would probably not use the CLT in her classroom.

Elementary Battery (Form W, Grades 4-6)

The two teachers who evaluated the Elementary Battery of the CLT differed substantially in their opinions. A comparison of their views provides an outline of the test's chief characteristics.

Teacher A felt that one of the test's major strengths is that it makes possible a comparison of the abilities of local students with those of students throughout the nation. Teacher B apparently felt that the test has no great strengths, since none were mentioned in her response. Both teachers agreed that the test is marred by several weaknesses. Both mentioned the spelling test, which requires the student to select one incorrectly spelled word from a list of four, but not to spell it correctly. Teacher B considered this kind of test "unrealistic" and Teacher A suggested that "a dictated spelling test might provide a better measure of ability." Both criticized Sections A and B (Capitalization and Punctuation), but for different reasons. Teacher B objected to the numbers placed within the sentences because they tend to break the pattern of thought, thereby confusing the slower, less confident child. Teacher A felt that pupils should be required to locate the places where capitalization and punctuation are required, and that the suggestion of possible choices invalidates pupil responses.

While Teacher A seemed to think that the test adequately measures what it claims to measure, covers material basic to the English curriculum, and is related to informal standard English, Teacher B judged it inadequate on these three points. She considered the test too limited and too simple to allow students to "show what they really know." While both teachers agreed that the test is neither long nor comprehensive enough to provide a reliable measure of achievement, they differed again in their opinions of the norms. Teacher A felt the stated norms would measure student performance accurately, while Teacher B felt that her students had rated too high in the past (she had administered the test before the norms were updated).

Finally, Teacher A suggested that the data presented in the supplementary materials were satisfactory, but her stated objections to the test itself would prevent her from using it except to diagnose general strengths and weaknesses and to compare the progress of students within a class.

Junior High Battery (Form W, Grades 7-9)

The teacher who evaluated this section of the CLT seemed to have formed a favorable opinion. He responded affirmatively to every question, suggesting only that the punctuation section would be more reliable if items on use of the colon and semicolon were included. He suspected that the stated norms might not apply to his students, many of whom come from non-English-speaking homes and have trouble with spelling and usage. His only other objection to the test concerned items 88-99, which test the ability to identify complete sentences. These, he felt, could possibly cause some confusion for students who have not made a distinction between grammatically complete sentences and complete thoughts. In general, however, this teacher approved the CLT and stated that he would like to use it in his classroom. He listed its major strengths as its ease of administration and its success in testing what it claims to test.

Advanced Battery (Form W, Grades 9-14)

The teacher who evaluated the Advanced Battery of the CLT gave it a strongly unfavorable review. He appraised Parts A and B of Test 5 as excellent standard measures of capitalization and punctuation and rated Test 6 as adequate for testing spelling mastery. But his objections to Test 5, Section C (Word Usage) caused him to respond negatively to all other questions except one. He remarked that the Word Usage Test "seems to fall short of the approaches to the new grammars. Terminology is too limited to a traditional approach and might be misleading to students." He especially objected to items such as number 85: "There are *[1]eight [2]five* different parts of speech." Although this section is entitled "Word Usage" and instructs students to choose "the correct or better word" in each item, 29 of the 48 items require students to know and apply traditional grammatical rules and definitions. Furthermore, the 19 items actually concerned with word usage frequently include distinctions which are rapidly breaking down — e.g., those between *lie* and *lay*, *sit* and *set*. The only question to which this teacher responded affirmatively concerned the CLT's supplementary materials, which he considered clear and comprehensive.

IV. Published Reviews

Richard E. Schutz, Professor of Education and Director of the Testing Service at Arizona State University, Tempe, Arizona, rates

the CLT unsatisfactory as a language test. His review concentrates on the sampling techniques used to obtain the norms and on the norms themselves. On the whole, he considers the 1963 standardization program so ill-defined that "it is impossible to separate sampling error from true variability in assessing any of the normative differences [between the 1957 and 1963 figures]" (p. 545). He judges the norms sample (15,351 students) inadequate and suggests that more information be made available regarding the number of schools and states involved in the sample, the method of obtaining it, and the differences between the 1957 and 1963 figures.

Like the teacher who evaluated the Advanced Battery, this reviewer questions the appropriateness of many of the items in the Word Usage Section. He admits that "the lag between scientific advances and classroom instruction is probably sufficient to maintain the curricular validity of 'usage' items for the majority of classrooms for some time to come" (p. 546). Nevertheless, he considers the title "Language Test" unsuitable to a test which does not include such topics as dialect differences, structural patterns, and verbal expression.

Cooperative English Tests, 1960 Revision

I. General Information

Grades 9-12, 13-14; 1940-60; 6 scores: vocabulary, reading comprehension (level, speed, total), English expression, total; Forms A, B, C; two levels; two tests (reading comprehension and English expression) available in separate booklets or a single booklet; directions, manual, technical report available; separate answer sheets required; $4.00 per 20 copies of either test; $6.00 per 20 tests (single booklet); 40 (45) minutes per test; revision by Clarence Derrick, David P. Harris, and Biron Walker; Cooperative Test Division.

The manual claims that the CET measures "achievements of high school and college students in two fundamental English areas: reading and written expression." No grade designations appear on the tests, so advanced or slow students can be given the next level (higher or lower) if the examiner wishes. As all forms of the test share the same general directions and time limits, different forms can be administered at the same time. Converted scores for all forms of the test are on the same scale, so that student scores are directly comparable.

Proceeding on the assumption that vocabulary is the best single index of verbal skill, the publishers have included a long and carefully worked-out vocabulary test. The English Expression Test is divided into two parts: Part I, (Effectiveness), requiring a choice of the most precise definition; and Part II, (Mechanics), including usage, spelling, punctuation, and capitalization. Items in Part II are designed to stimulate the proofreading process. The student's score on the Expression Test is intended "to describe [his] ability to select appropriate usages and see incorrect usages. It is not a direct measure of writing ability, but evidence suggests that ability to do well on this kind of test is related to ability to write well in an essay situation" (p. 7).

The publisher suggests several ways in which individual and group scores can be used and provides detailed instructions for interpreting scores. The manual explains many terms which often prove troublesome to those unfamiliar with testing and scoring methods (e.g., "percentile rank," "norms table," etc.).

II. Use in Wisconsin

The CET is used by 11, or 4 percent, of the 306 respondents to the first questionnaire. Eight, or 72 percent of the 11 users, rated the test satisfactory because they considered it comprehensive but not too long, and because scores coincided with scores on other measurements. Three, or 27 percent, found it unsatisfactory because it is not comprehensive enough and does not allow students to correct the errors they spot.

III. Teacher Evaluation

The teacher who evaluated form 2A of the CET seemed to have gained a favorable impression. Although she felt that Part II of the Expression Test (Mechanics) contains "too many obvious or gross errors in usage to be practical," she commended this section for its general "breadth and scope." She answered all other questions affirmatively and found some phases of the supplementary materials most useful for reteaching, or individualized teaching.

At the time she responded to our questionnaire, this teacher was using the Expression Test in her classroom and was apparently well satisfied with it.

IV. Published Reviews

All three of the following reviewers commend the CET English Expression Test and agree that it comes as close as an objective test can to measuring writing skill accurately.

Leonard S. Feldt, Professor of Education at the State University of Iowa in Iowa City, Iowa, praises the authors for basing the content upon a study of the frequency of student errors in actual themes. (See the Manual, p. 20.) He notes that the publishers state that the ability to organize ideas, to break a composition into paragraphs, and to select phraseology more appropriate to one kind of writing than to another is not measured by the CET. Both this reviewer and the publishers suggest that those teachers interested in measuring such abilities give serious consideration to the *Sequential Tests of Educational Progress*, Essay and Writing Tests.

Mr. Feldt concludes by suggesting that certain additions be made to the norms data. He commends the wealth of technical data provided on validity, reliability, scaling, and norming, but wonders

why reliability data are provided for grades 10-12 only and why the norms are not more complete. He also questions the publisher's apparent lack of concern about the unreliability of the total scores of individual students.

Margaret F. Lorimer, Associate Professor at the Office of Institutional Research at Michigan State University, East Lansing, Michigan, would agree with Mr. Feldt that the CET norms could be more complete. She states that the high school norms are "hardly representative of the various regions or of the general population" and that the high schools in the sample "are located for the most part in small towns in rural areas" (p. 554).

Miss Lorimer feels that the inclusion of spelling errors in the "Mechanics" section lessens the test's diagnostic value. She objects also to the superficiality of the usage items. However, she approves the introduction of a new type of mechanics item which requires students to find as well as to correct errors, and grants that the "Effectiveness" items probably come as close as possible to measuring a student's ability to use words precisely.

John C. Sherwood, Professor of English at the University of Oregon at Eugene, also raises some objections to the CET but nevertheless considers it one of the best objective tests available. He criticizes the "Effectiveness" section for devoting 20 out of 30 items to exact word choice and for leaving only 10 items to cover all other stylistic problems. He also notes that in several diction items, because of contextual ambiguity, more than one answer could be considered correct.

Mr. Sherwood goes on to praise the authors and publishers for the "formidable effort that went into preparing both the test and the technical apparatus that goes with it" (p. 557). He concludes that the test is generally efficient, brief, and comprehensive, and that it includes items of a relatively high quality. While he doubts that a liberal grammarian would give unqualified approval to the more conservative usage items, Mr. Sherwood does believe that the CET items test the kind of expression that occurs in the ordinary writing process, and he recommends that the test remain in use.

Differential Aptitude Tests

I. General Information

Grades 8-13; 1947-63; 9 scores including language usage (spelling, sentences); Forms A and B ('47), L and M ('62); Manual ('59); individual report forms for all forms; individual report folder for Forms A and B; casebook available; separate answer sheets required; language usage test (Form A or B) available separately at $3.00 per 25 tests; 35 (45) minutes; George K. Bennet, Harold G. Seashore, and Alexander G. Wesman; The Psychological Corporation.

The DAT specimen set does not include a statement of the test's development and objectives. Directions for Administration and Scoring (3rd Edition) are included and contain a description of the test materials, lengthy instructions to test administrators, scoring information, and a comprehensive description of the norms and profiles (including norm tables for both boys and girls at each grade level).

II. Use in Wisconsin

The DAT language usage test is used by six, or 2 percent, of the respondents to the first questionnaire. All rated the test satisfactory and praised its ease of administration and scoring, objectivity, and brevity. One user mentioned that he found it useful in grouping students; another noted that it helped his school set up its own testing program.

III. Teacher Evaluations

The teacher who evaluated Form A (1947) of the DAT language usage test responded affirmatively to only one question: he considered the included material a valid part of the English curriculum. He strongly objected, however, to the content of most of the items and suggested that the Manual include a description of the method of selecting the words on the spelling list, and that the form of Part I (Spelling) be altered. At present, students are given a list of 100 words and told to mark whether each is spelled correctly or incorrectly. They are not asked to spell misspelled words correctly.

This teacher stated that he would prefer several spellings of the word, perhaps three incorrect and one correct. He also suggested that the words be presented in the context of a sentence and predicted that the bizarre spellings of some words ("consinment," "relize," etc.) would make them unrecognizable to students.

In Part II (Sentences) this evaluator cited several items such as, "it is me" and "got hurt," which students would mark wrong if they did not accept informal English usage.

Finally, the teacher criticized the use of two scoring keys, one for Right and one for Wrong answers. He felt that hand scoring by this method was unnecessarily tedious. Because of his many objections to the test, he stated that he would not use it in his classroom.

IV. Published Reviews

[Note: The following reviews are discussions of Forms L and M (1962).]

J. A. Keats, Reader in Psychology at the University of Queensland, Brisbane, Australia, begins his review by discussing the revisions made in the 1962 forms of the DAT. The content of the Spelling Test is the same; however, the revised form of the "Sentences" section is entitled "Grammar" and contains ten additional items with only one correct response per item. (Each item in the 1947 forms was divided into five parts: each part could have contained an error.) Where scores on the 1947 Edition were corrected for guessing, scores on the new edition are based upon the number of correct responses; the reviewer commends this change.

Mr. Keats notes, however, that out-of-date standards of usage of the "Grammar" section still have not been altered. His other suggestions pertain to technical matters such as inclusion of multiple correlations of validity studies, outlining of the research methods used in establishing percentiles as the basis of the norms, etc. In summary, he commends the changes made in the 1962 Edition of the DAT, but feels that more changes and evidence for the changes made are necessary "to enable the battery to represent the standard to which others should aspire" (p. 1005).

Richard E. Schutz, Professor of Education and Director of the Testing Service at Arizona State University, Tempe, Arizona, devotes his review of the DAT to a discussion of improvements made

in the test's technical apparatus. He notes that the norms sample (50,000 students from 195 schools in 43 states) adequately represents the U. S. population with respect to geographical distribution and community size. The norms are relevant for fall testing programs, but spring norms must be obtained by interpolating between successive grades tested in the fall.

Mr. Schutz notes that certain criticisms of the 1947 Edition of the DAT have been met by the following changes: alteration of scoring so that only right answers are counted; correlations with other tests now provided; and the interpretation of results enhanced by the addition of a report folder. Other deficiencies, however, have not been remedied: no information concerning item analysis is given, for instance, nor has anything been done to correct the apparent duplication of material in various subtests.

These criticisms and suggestions pertain to the DAT as a whole, but may be helpful to those concerned only with the Language Usage Test. Mr. Schutz's overall evaluation is mildly positive.

Essentials of English Tests, Revised Edition

I. General Information

Grades 7-13; 1939-61; 6 scores: spelling, grammatical usage, word usage, sentence structure, punctuation and capitalization, total; Forms A, B ('61, identical with 1939 and 1940 forms except for revisions in 12 items); Manual ('61, essentially identical to '44 Manual except for wording changes); reliability data and norms same as published in 1939-44; $2.50 per 25 tests; 45 (50) minutes; original edition by Dora V. Smith and Constance M. McCullough; revision by Carolyn P. Greene; American Guidance Service.

The preface to the Manual of Directions explains that the EE was revised to permit it "to keep pace with current restudy and evaluation of the English language in terms of the ways in which people speak and write."

The manual continues by describing the test, stressing that each area is tested in context and that students are required to correct the errors they spot. The publishers claim that the test may be used as a survey test for observation of "the variety of English abilities represented in a given class, school, or system as a whole." They further maintain that "the chief value of the examination probably lies in its diagnosis of individual strengths and deficiencies in the English abilities tested." Item validity is said to rest upon studies of frequency of use and error, frequency of appearance on English placement examinations administered by 130 colleges and universities, and "universal agreement among English authorities."

The publishers state that they "are more concerned that teachers interest themselves in the performance of individual pupils than in any group comparisons." The EE is designed, therefore, to help the teachers group students and plan remedial teaching programs. For teachers who wish to compare their students to a national sample, however, norms "based on the performance of 36,480 pupils of grades 7-12 in all sections of the country" are reported in terms of percentile scores by grades. Norms are based on mid-year administration.

II. Use in Wisconsin

Sixteen, or 5 percent of those respondents to the first questionnaire who employ standardized tests use the EE. Fourteen, or 87.5 percent, rated it satisfactory, mentioning its helpfulness in grouping students (one of the stated objectives), its comprehensiveness, and its diagnostic value. One user who rated the test satisfactory qualified his judgment by calling the spelling section "weak" and the word usage section "narrow in range." One user rated the test unsatisfactory because of the formal English usage it espouses, and one did not rate it.

III. Teacher Evaluations

Both teachers who evaluated the EE formed favorable opinions, with slight reservations. While Teacher A felt that it would not thoroughly measure the abilities of his students, Teacher B judged it "quite inclusive." She considered the Sentence Structure Section especially valuable. Both teachers responded affirmatively to questions two and three, but differed in their answers to question four. Teacher A responded affirmatively, but Teacher B judged the usage sections of the test "unrealistic" and unrelated to informal standard English. She commented: "When industrial and political leaders, school administrators and teachers consistently make many of these errors, it is rather hard to convince students that 'correct' usage has much validity in their lives." She considered the test reliable, however, while Teacher A questioned its reliability for students at advanced grade levels.

Neither teacher felt that the stated norms would provide a realistic guide for measuring student performance. Teacher A seemed to imply that his students would rate too high, while Teacher B felt that her students would rate low because of the substandard English spoken in their homes. Neither teacher presently uses the EE, but despite their respective reservations, both stated that they would like to try it.

IV. Published Reviews

J. Raymond Gerberich, Visiting Professor of Education at the University of Maryland, College Park, Maryland, reviews the EE unfavorably. He cites several discrepancies that careful editing would have eliminated: for instance, some items in Part IV (Sentence Structure) fail to include the same details in all

four options; one item presents one good and three bad options; and another item offers three acceptable sentences and one unacceptable one rather than the opposite. Mr. Gerberich objects also to the methods of scoring Part III (Word Usage) and Part V (Punctuation and Capitalization). He feels that further instructions for scoring errorless sentences and for handling scores containing fractions in Part III should be included, and states that the arrangements of points in Part V makes objective scoring almost impossible.

Turning to the technical apparatus, Mr. Gerberich notes that validity, reliability, and comparability of results receive very sketchy attention in the manual and norm tables. No norms are given for grade 13. The 1940 and 1961 norms are identical, which suggests that both must be based on figures obtained before 1940.

This reviewer concludes that the 1961 revision and its 1939-40 predecessor differ insignificantly in content, and not at all in accompanying norms or evidence concerning reliability and validity. He also notes that the revision fails to incorporate the suggestions for improvement made in *The Third Mental Measurements Yearbook* (1949), the recommendations for authors and publishers of achievement tests published in the 1950's, or the technical recommendations. For these reasons, Mr. Gerberich does not recommend the EE to teachers of English.

Greene-Stapp Language Abilities Test

I. General Information

Grades 9-13; 1952-54; 5 scores: capitalization, spelling, sentence structure, punctuation, usage; Forms AM ('52), BM ('53); Manual ('54); $6.40 per 35 tests; $1.75 per 35 IBM answer sheets; 80 (95) minutes in two sessions; Harry A. Greene and Helen S. Stapp; Harcourt, Brace & World, Inc.

In the specimen set the GS is advertised as a "comprehensive measure" of proficiency in the use of the English language and a "reliable guide" for individual instruction. National percentile norms based on administration to 8,415 students in 26 high schools from 15 states are provided by grade for grades 9-12, for each subtest and for the total score. The test is allegedly designed for ease of administration and scoring; the manual includes extensive instructions for interpretation and use of test results. Content and methods of testing are standard, with two exceptions: in Test III (Sentence Structure and Applied Grammar), the student is asked to choose the statement which tells what should be done to improve incorrect sentences; in Test V (Usage) he is told to choose the statement which tells why an incorrect word in a sentence is wrong. In other words, he is required to spot errors and in one case to choose between given methods of correcting them; but he is not asked to rewrite sentences himself or to substitute appropriate for "incorrect" words.

II. Use in Wisconsin

Of those respondents to the first questionnaire who employ standard tests, eight, or 3 percent, use the GS. Five, or 62.5 percent, rated it satisfactory; one of these noted that it is concerned with "rhetoric and fine discriminations" in sentence structure. It was praised also for its comprehensiveness and helpfulness to teachers willing to analyze and follow up the results. One test user considered it too involved and technical on grammar to be satisfactory. Two, or 25 percent, did not rate it or comment upon it.

III. Teacher Evaluations

The opinions of the two teachers who evaluated Form AM of the GS differ considerably and will be discussed separately.

The first teacher accepted the content, norms, and supplementary materials, and considered the test a success in measuring what it claims to measure. However, she criticized its neglect of certain established conventions of punctuation and its failure to accept changing usage in a few test questions. She felt that many possible responses do not conform to informal standard English, and judged the test not comprehensive enough to measure the achievement of her students. She stated that she would not use this test in her classroom.

The second evaluation of the GS represents the opinions of three teachers of "upper level" sophomores at the same school. It was phrased as the statement of one person to indicate that the three were in agreement. Thus, it will be treated as a single evaluation.

The teachers formed a generally favorable opinion of the test. They considered certain portions of Test IV (Punctuation) "outdated" — e.g., "the use of the comma in restrictive and nonrestrictive clauses, as well as before 'and' in a series." They criticized the ambiguity of some of the choices in Test V (Usage and Applied Grammar) and objected to the use of unfamiliar terms such as "copulative verb." It was felt that both of these factors might lead a student to select the wrong answer even though he knows the correct one. The teachers suspected that their "upper level" students, "who seem to speak and write correctly by instinct," might not do well on those sections (Tests III and V) which require the citing of "rules" to explain why sentences or words are incorrect.

Despite these objections, however, the teachers praised the test for several reasons. They considered it 'easy to administer' and successful in testing student ability to recognize correct forms. They especially liked the capitalization test and the form of the spelling test, in which the repetition of words helps to test whether a student actually recognizes the correct or incorrect forms or whether he merely makes an accurate guess. They accepted the test's relation to "informal standard English," reliability, and supplementary materials, and stated that they would like to try it with their students.

IV. Published Reviews

The following reviews may be found in *The Fifth Mental Measurements Yearbook*, Oscar K. Buros, ed.

Richard A. Meade, Professor of Education at the University of Virginia, Charlottesville, Virginia, considers the GS an adequate measure of the skills it includes. He finds the manual "adequate" (p. 345), the directions for scoring comprehensible, and the instructions for interpreting results and devising remedial work clear and useful. In his opinion the capitalization, spelling, and punctuation subtests are "adequate and . . . geared to actual performance at this level" (p. 345). In the usage and sentence structure sections, however, Mr. Meade discovers "more stress on grammatical understanding than on ability to identify correct or incorrect structure and usage" (p. 345). As an example he cites the dependence of a high score upon knowledge of grammatical rules. Furthermore, he notes that the usage test apparently "takes no note of colloquial (informal) usage" (p. 345). According to his calculations, one-third of the "incorrect" usages are acceptable to many people for informal purposes. Students are not informed of the test's standards; thus those who do not consider formality of usage the basis of "correctness" may not score well.

By and large, Mr. Meade judges this a "well-constructed [test] which adequately covers the areas it includes" (p. 345). It is convenient and usable if the user allows for the weaknesses in the areas of usage and grammar.

Osmond E. Palmer, Associate Professor at the Office of Educational Services, Michigan State University, East Lansing, Michigan, would agree with Mr. Meade that the GS can be helpful if used properly. He finds the manual "unusually complete" for a test of this kind. In his opinion the subtests are long enough and the capitalization and punctuation sections are especially thorough. However, he questions the nature of the items used in the punctuation test. The answer sheet presents four possible punctuation marks to be considered in each situation, but in half of the cases the choice is reduced to a comma, a period, or nothing. In addition, Mr. Palmer objects to the format of the spelling test (four different words, three or four of which may be misspelled, are offered; students must decide which, if any, is correct). The strangeness of some of the misspellings and the absence of certain commonly misspelled words (arctic, separate, etc.) are also questioned. In the sentence structure and usage tests the reviewer finds many responses inapplicable to the items. Furthermore, many responses consist of statements of principle which are not true. Thus students who expect the statements to be either true or false

may be confused by having to consider their accuracy as well as their applicability.

Finally, Mr. Palmer suspects that the test may be speeded, which would alter the reliability and significance of scores. In his words, "the difference between two scores may be due to greater knowledge of the matters tested, or it may be due merely to speed" (p. 346). In a word, he believes that although the GS may be useful if used properly, other tests will probably prove more fruitful.

Iowa Tests of Basic Skills

I. General Information

Grades 3-9; 1955-56; 6 scores in language arts area: vocabulary, spelling, capitalization, punctuation, usage, total; Forms 1 ('55), 2 ('56), 4 ('64); Teacher's Manual ('64); Administrator's Manual ('64); Profile; Class Record Sheet; Pupil's Report Folder; IBM or MRC answer sheets must be used; 84¢ per test; see publisher's Standardized Tests and Scoring Service Catalog for prices of answer sheets, etc.; Vocabulary and Language tests require 84 minutes; E. F. Lindquist, A. N. Hieronymus, et al.; Houghton Mifflin Company.

The ITBS claims to be the only test battery that measures a pupil's ability to use his acquired skills. It also claims that the tests for each grade are *adapted specifically* to that grade and that *complete continuity of measurement is provided in grades 3-9*.

According to the publishers, the norms are "*really national in character*," representing all geographic regions and sizes of schools. The Administrator's Manual notes that two types of norms are provided: grade norms and percentile norms within a grade. The 1964 norms based on a 1963 national standardization program are now provided. The standardization program was carried out in cooperation with the authors and publishers of the *Lorge-Thorndike Intelligence Tests* and the *Tests of Academic Progress*. Detailed information concerning the obtaining of the norms sample is included in the Administrator's Manual.

The Administrator's Manual also discusses the nature and purpose of the tests, organization of a local testing program, interpretation of test scores, and use of test results to improve instruction. The Teacher's Manual provides directions for administration and scoring, tables of percentile norms, and suggestions for interpreting and using test results. An added feature is the Pupil's Report Folder, which explains the purpose of each test and provides space for plotting the student's profile.

The items in the Vocabulary Test consist of a word in context followed by four possible definitions. It is claimed that "the immediate purpose of each item is to determine if the pupil knows the

meanings of all the words used in the item. Thus, a 40-item vocabulary test may sample as many as two hundred words from his general vocabulary . . . ."

The Language Test is divided into four separate subtests: spelling, capitalization, punctuation, and usage. The basic type of item employed in all four language tests may be described as the "find-the-error" type. The authors believe that "this type of item most clearly differentiates between those who habitually use correct language and those who have not developed functional habits of correct language." In the development of content, "the authors have attempted to draw upon the best of current practice, as evidenced in courses of study, textbooks, and research studies." The spelling test items contain four words, one of which may be misspelled. The authors consider this item type superior to that which presents four possible spellings of the same word, and claim that it measures almost exactly the same skills that a dictation list test would measure. In the capitalization and punctuation tests the authors have included materials which might have been found in children's work. The items in both tests "include one or two sentences extending over three lines of approximately equal length. The student is instructed to identify the line which contains an error or to elect a fourth response indicating the total absence of any errors." This type of item was adopted after careful investigation and is said to be similar to the "free-response" type of item used in modern language tests. Again the "find-the-error" type of item is employed, and it is claimed that this type of item differentiates between those who merely *know* correct English and those who actually *use* it. As with the other language tests, specific studies of frequency of errors were consulted in designing the usage items. However, it is not indicated that this test attempts to measure students' knowledge of anything but formal English usage.

II. Use in Wisconsin

According to the results of the first questionnaire, the ITBS is the most widely used standardized test in the state. Of the 306 respondents who employ standard tests, 123, or 40 percent, use this one. Sixty-seven, or 54 percent, of these rated it satisfactory for the following reasons (in order of frequency):

1. It is diagnostic of individual strengths and weaknesses.
2. It is comprehensive.

3. Its results coincide with students' actual performance and teachers' evaluations.
4. Its norms are well-constructed; it provides a basis for annual comparison of student achievement.

Twenty, or 16 percent, judged it unsatisfactory for the following reasons (in order of frequency):

1. It measures mechanical skills only; it does not test the ability to write or to use skills functionally.
2. The spelling test is poor.
3. It encourages guessing by testing what children will be taught rather than what they have learned. [Note: compare this with the publisher's claim that this "is the only battery that measures the pupil's ability to use his acquired skills. No test . . . is concerned with repetition . . . of formal facts or rules."]
4. It does not test over-capitalization and over-punctuation.

Thirty-six, or 29 per cent of the test users neither rated it nor commented upon it.

Teachers in the Milwaukee school system, where the ITBS is universally employed, were questioned about it in May, 1964. They rated the four skills tests (including the Language Test) adequate in terms of item construction, content, relevance to the curriculum, reliability, usability of results, and standardization. Only four of the 101 teachers who returned the questionnaires felt that a change to a new test was warranted. Ninety-two were satisfied that it does adequately measure the academic ability of pupils. In December, 1964, Milwaukee guidance directors were also asked to estimate the value of the ITBS for purposes of guidance and educational planning. Ten found it "essential"; three rated it "quite helpful"; one considered it of "average" value; and none felt it had "little" or "no value."

III. Teacher Evaluations

Both teachers who evaluated Form 4 of the ITBS formed highly favorable opinions. In fact, both responded affirmatively to all questions on the inquiry. There were two qualifications: one teacher suspected that her slower students might guess at many of the items and thus achieve higher scores than they deserved; the other felt that her more creative pupils might not do as well as they should on an objective test.

Both teachers considered the test extremely comprehensive. One judged it "easy to administer" and "understandable to children." The other praised its ease of scoring and especially liked the convenience of the single booklet edition and the fact that certain items have been assigned to particular grade levels. Both teachers had used this test in their classrooms and stated that they will continue to do so.

· Although only two teachers were asked to evaluate the ITBS, a third sent us the following comments of her own accord. They are pertinent to the Language Test and contain certain criticisms which run counter to the publisher's claims and to the other teachers' opinions. This teacher examined the usage items carefully and concluded that in the light of modern research, only one or two of these might be questioned. She felt, however, that achievement of a high score might be too closely dependent upon reading skill. To remedy this she suggested adoption of shorter sentences concerned with "school experience common to all pupils," which "might serve to focus attention on usage." Finally, she suspected that the "find-the-error" type of item acclaimed by the publisher transforms language activity "from an expressional act to a recognition act" and encourages "a restrictive approach to teaching communication skills."

IV. Published Reviews

The following reviews are taken from *The Fifth Mental Measurements Yearbook,* Oscar K. Buros, ed.

Vergil E. Herrick, Professor of Education at the University of Wisconsin, Madison, Wisconsin, points out that the ITBS "cannot be considered an achievement battery in the usual sense of measuring knowledge in the common content areas of the elementary school curriculum . . . . The focus of these tests is on the evaluation . . . generalized intellectual skills and abilities . . . ,not on content achievement per se" (p. 31). He notes that the publishers consider measurement of these skills more valuable in "the improvement and individualization of instruction and educational guidance" than measurement of specific knowledge, but he contends that both kinds of measurement "are necessary to proper educational evaluation" (p. 31).

Mr. Herrick praises the authors of the ITBS for the following achievements: the continuity of measurement attained by the sin-

gle booklet; reliabilities "sufficiently high for individual diagnosis and prediction" (p. 31); the curricular validity of the items; the thoroughness of curricular analyses designed to help teachers plan remedial instruction; the development of norms for performance at the beginning, middle, and end of the year; and the comprehensive standardization sample. However, he notes that the intercorrelations of subtests indicates "a heavy loading of all subtests with vocabulary and reading skills" (p. 32). [Note: compare this comment with the third teacher evaluation, above.]

Moreover, Mr. Herrick objects to certain important features of the Vocabulary and Language Test. Although he considers the vocabulary sample more adequate than that employed in many similar tests, he still describes it as "limited." He notes the claim that knowledge of response words as well as of stimulus words is checked, but suspects that this may be invalidated by the difficulty of many of the response words. His strongest criticism of the Vocabulary Test is that it devotes "little attention . . . to the evaluation of tools involved in word recognition and verification." [Note: He does not define these "tools."] In his opinion the Vocabulary Test is "more a test of experiential background or intelligence than of basic skills" (p. 33).

Mr. Herrick's next criticism echoes that of the third teacher (above): that the use of "find-the-error" items "tends to emphasize the editorial aspect of language use and not the dynamic, functional, creative aspect which exists when one writes" (p. 33). He considers the language subtests well-constructed and valid in relation to language arts texts and research studies, but questions whether "certain common and persistently used language skills" are covered adequately. He does not enumerate these skills, but presumably means those related to the "dynamic, functional, creative aspect" of language: employment of various sentence patterns, ability to organize coherent paragraphs and to choose words which precisely convey intended meaning, etc. Despite his objections, however, Mr. Herrick concludes that the "curricular validity, careful construction, . . . adequate norms . . . , and high reliabilities" (p. 33) of the ITBS classify it among the best available at this time.

G. A. V. Morgan, Senior Psychologist at the North Wales Child Guidance Clinics, Denbighshire, Wales, praises the technical achievements of the ITBS, but wonders whether too high a price

has been paid for its technical excellence. Most of his criticisms of the battery pertain to the Language Test. Like Mr. Herrick, he considers the content "limited" and concludes that "the demands of item form seem . . . to restrict considerably the flexibility and range of a language attainment test" (p. 34). He considers the test "overweighted" with capitalization and punctuation. He objects to these sections on the grounds that they do not "fully measure the control of expression and sense of sentence structure which punctuation exists to serve." The same objections apply to the usage test, which Mr. Morgan considers "limited in range and . . . 'textbookish' in character." In short, he feels that the Language Test may possibly encourage a restricted approach in teaching, sacrificing too much to the demands of objective form and rapid scoring. Although Mr. Morgan accepts the ITBS as the best of its kind and a model of test construction, one suspects that he would not strongly recommend it to teachers of English.

Iowa Tests of Educational Development

I. General Information

Grades 9-12; 1942-63; 2 scores pertinent to language: correctness and appropriateness of expression, general vocabulary; IBM and MRC; 2 editions (single booklet and separate booklet); separate answer sheets must be used; Forms X-4, Y-4; examiner's manual; administrator's manual; teachers' and counselors' manual; student profile sheet; single booklet edition rented only; fee: $1.25 per student (including scoring service); separate booklet edition may be purchased at $2.40 per 20 tests; Test 3 requires 60 minutes, Test 8, 22 minutes; prepared under the direction of E. F. Lindquist and Leonard S. Feldt; Science Research Associates, Inc.

According to the manual for teachers and counselors, the ITED is "designed to provide a comprehensive and dependable description of the general educational development of the high school pupil" (p. 6). The tests have two major purposes: to keep teachers and counselors "acquainted with the educational development of each . . . pupil"; and "to provide the school administrator with a more dependable and objective basis for evaluating the total educational offering of the school" (p. 7).

This survey is concerned with two of the ITED's nine subtests: Test 3 (Correctness and Appropriateness of Expression) and Test 8 (General Vocabulary). Test 3 is intended to measure "some of the basic elements in correct and effective writing: punctuation, usage, capitalization, spelling, diction, phraseology, and organization" (p. 15). With the exception of spelling, these are not tested separately as in other tests. Instead, the student is given a letter and three passages designed to resemble the writing of a high school student. He must decide whether each underlined portion is acceptable, and, if not, which of the alternative forms given in the right-hand column is appropriate. It is claimed that this kind of test "parallels closely the task which the pupil faces in an actual, writing situation" (p. 16), and thus measures his ability to apply his knowledge of language. "Usages and practices on which there is not substantial agreement among English teachers" (p. 16), as well as elementary skills which most high school students have mastered, are not included. The spelling test is of the standard

type: 15 groups of four words each are presented and the student is asked to choose which word (if any) is misspelled in each group. The publishers stress that Test 3 does not attempt to measure "the more subtle or intangible elements of composition ability" (p. 16), and recommend supplementation of scores by performances on actual compositions.

The supplementary materials accompanying the ITED seem inclusive and comprehensible. The manual for teachers and counselors includes discussions of pupil scores and profiles, instructions for interpretation and use of test results, and tables of national and percentile norms. The examiner's manual contains instructions for administering the tests; the administrator's manual consists of directions for administration and follow-up of the local testing program and statistical data on reliability, validity, and standardization.

II. Use in Wisconsin

The ITED is used by 61, or 20 per cent, of the respondents to the first questionnaire who use standard tests. Thirty-five, or 57 percent, rated it satisfactory for the following reasons in order of frequency:

1. It measures achievement accurately.
2. It is useful for grouping students.
3. It provides a general view of student progress.

One respondent commented, "It measures ability to think rather than mere factual recall." Ten respondents (17 percent) judged it unsatisfactory for these reasons, again in order of frequency:

1. Test 3 is conservative in its emphasis on mechanics and grammar.
2. Test content is inadequate.
3. Content is too general.
4. No state norms are provided.

Sixteen, or 27 percent of the test users did not rate it or comment upon it.

III. Teacher Evaluations

Neither of the two teachers who evaluated Form X-4 of the ITED formed a favorable opinion of it. Their criticisms, however, are different enough to warrant separate presentation.

In the opinion of one teacher, Test 3 (Correctness and Appropriateness of Expression) has three major weaknesses. The first of these is "imbalance." Fifty per cent of the test, she objects, "is devoted to items of diction . . . . Many of these [items] deal with infelicities and trivial, debatable points of diction." Second, she argues that "the assumption that the test should not cover certain elementary skills because they have been mastered by almost all high school students is not borne out by classroom experience." Finally, she fears that the "mixture of a problem in diction and a grammatical error within one situation or item might be confusing to students."

Furthermore, this teacher argues that the ITED fails to measure those abilities it claims to measure. She denies that "the test parallels closely the task which a pupil faces in an actual writing situation" (T. and C. Manual, p. 16), objecting that it "lacks sufficient material to test some of the basic elements such as usage, capitalization, and organization." She suggests that more items deal with "errors in verbs and pronouns — the greatest source of usage errors." In her opinion the test emphasizes formal, rather than informal, standard English and employs vocabulary which is often too "sophisticated" for ninth grade students. She considers the norms adequate in relation to content, but contends that much of the content is "concerned with minor points, too vague and variable to be of value in diagnosis and remediation". She seems to believe that probably a test prepared by English experts would be more useful to classroom teachers.

The second teacher grants that Test 3 covers many important phases of grammar but fears that the high degree of subjectivity which characterizes the item form will be "confusing" to students. She also considers the entire test too long to be properly administered within an hour. She agrees with the first teacher that the test does not measure writing ability, but rather the ability "to proofread *another's* method of expression." Although she accepts the content as a valid part of the English curriculum, she does not judge the method of testing adequate, nor does she regard the emphasis upon the more formal aspects of written English conducive to the development of standards of informal speech. In other words, this teacher considers Test 3 of the ITED too limited in content and restrictive in approach, and concludes that she would not use it in her classroom.

IV. Published Reviews

Ellis Batten Page, Professor of Education and Director of the Bureau of Education, Research, and Service at the University of Connecticut, Storrs, Connecticut, devotes his review to a consideration of the technical and statistical merit of the ITED. He notes the publisher's suggestion that each administrator evaluate the test content in order to determine "an individual kind of face validity" (p. 49). This kind of evaluation, he contends, ought to be supplemented by statistical proof of validity to reduce the possibility of judgmental error. Nor is correlation of test scores with high school and college grades an indication of validity, since grades may be an inferior indication of achievement. The reviewer commends the publishers for taking note of this.

Mr. Page objects, however, that students who guess may do better than those who do not. He believes that a computer could probably be used to reduce the occurrence of obviously patterned responses. He goes on to discuss the population sample, which he considers "most respectable" (p. 51). He commends the inclusion of different kinds of norms, so that an administrator may compare his schools with a national population of students or with one school.

In short, Mr. Page comments upon the ITED from the administrator's point of view. From this perspective he considers it a well-constructed, efficiently scored, and comprehensively standardized measurement. He does not evaluate content, however, nor does he comment specifically upon the sections pertinent to the language arts. It is, therefore, impossible to predict whether or not he would recommend the test to teachers of English.

Alexander G. Wesman, Associate Director of the Test Division of the Psychological Corporation, New York, New York, reviews the announced goals of the ITED, which are not curriculum-oriented, but "emphasize ultimate and lasting outcomes of the whole program of education" (p. 51). He notes that the full length version requires eight hours and the class period version five and one-half, and feels that "the required investment of pupil and school time . . . make(s) it mandatory to consider whether or not there is adequate return for the expenditure involved" (p. 52). He suggests that before adopting the test, the administrator should ask: Is enough useful information provided for direct improvement of the pupil's education? Could equal information be obtained in

less time, or more useful information in the same amount of time? Is the change in abilities from year to year large enough to justify administration of the test every year?

While Mr. Wesman states that many of the items "call for the ability to generalize, to apply in new situations what has previously been learned in other settings, and to derive information from newly presented materials," he considers the spelling section of Test 3 and all of Test 8 "conventional measures of these fields of knowledge" (p. 52). He judges the SRA scoring and reporting services attractive and time-saving, but objects to the publisher's "lack of restraint . . . in putting forth claims . . . (which) are sometimes inconsistent with each other . . ." (p. 52). For instance, it is recommended that test results be used as a guide for curriculum revision, even though the test is "not constructed on the basis of an analysis of any specific high school courses" (p. 52).

Mr. Wesman also objects strongly to the test's lack of statistical data. In his opinion, "this kind of failure to present relevant data, even when these have clearly been available, typifies the program" (p. 55). He suggests inclusion of "validity coefficients for each of the tests against appropriate criteria," test-retest data from successive administrations, tables of intercorrelation with other tests, and evidence that sufficient growth occurs each year to warrant annual administration (p. 55). In short, he considers yearly retesting "wasteful" of time and money, and suggests that these resources "might better be devoted to testing for other abilities . . . which will yield new and useful information" (p. 55).

Metropolitan Achievement Tests

I. General Information

Grades 1.5, 2, 3-4, 5-6, 7-9, 9-12; 1931-64; IBM and MRC for grades 5-12; language subtest for grades 9-12 available separately; interpretive manual; individual profile and profile directions for *a-e;* profile for *f;* cumulative record card for *a-e;* Walter N. Durost, et. al., Harcourt, Brace & World, Inc.

a) Primary I Battery. Grade 1.5; 1931-62; 2 scores pertinent to language arts: Word Knowledge, Word Discrimination; Forms A ('60), B ('59), C ('61); directions for administering; $6.25 per 35 tests; language sections require 27 minutes.

b) Primary II Battery. Grade 2: 1932-62; 3 scores pertinent to language: same as *a* plus Spelling; Forms same as for *a*; directions for administering; $8.00 per 35 tests; language sections require 40 minutes.

c) Elementary Battery. Grades 3-4; 1932-62; 6 scores pertinent to language arts: same as *b* plus Language (Usage, Punctuation and Capitalization, Total); Forms same as *a* and *b*, plus D ('62); directions for administering; $8.00 per 35 tests; language sections require 69 minutes.

d) Intermediate Battery (Partial). Grades 5-6; 1932-62; Complete Battery including Social Studies Information and Science scores also available; 6 scores pertinent to language arts: same as *c* minus Word Discrimination and plus Parts of Speech under Language; 2 editions (hand and machine scored); directions for administering for each edition; Forms same as for *c*; separate answer sheets required; $9.80 per 35 tests; language sections require 57 minutes.

e) Advanced Battery (Partial). Grades 7-9; 1932-62; Complete Battery including Social Studies Information and Science also available; 2 editions (hand or machine scorable); 7 scores pertinent to language: same as *d* plus Kinds of Sentences under Language; directions for administering; Forms same as for *c*; separate answer sheets required; $9.80 per 35 tests; language sections require 76 minutes.

f) High School Battery. Grades 9-12; 1962-64; 2 scores pertinent to language arts: Spelling, Language; Forms AM ('62), BM ('63); directions for administering; norms booklet; separate answer sheets required; $10.50 per 35 tests; language and spelling tests require 42 minutes (plus 6 minutes for distributing materials, etc.).

In contrast with the non-curriculum-oriented achievement batteries, the MAT is designed to "measure what the schools are teaching" through "thorough analysis of current courses of study and instructional materials." The authors have grouped material into subtests "which make possible a more refined analysis of pupil competence" and which "are arranged in convenient work units." Distinctive colors are used to identify the materials for each battery and directions and scoring devices are designed for maximum efficiency. In addition, test scores are presented in conventional grade equivalents, percentiles, or stanines, and sufficient aids for constructive use of results are provided.

II. Use in Wisconsin

Of these respondents to the first questionnaire who use standardized tests, 45, or 15 percent, employ the MAT. Twenty-seven, or 60 percent, rated the battery satisfactory, primarily because it provides a basis for comparing students with local and national groups. Other strong points were listed as: correlation of scores with teachers' ranking; the continuous nature of the program; comprehensiveness; and diagnostic performance. Nine, or 20 percent, of the test users considered it unsatisfactory for several reasons: failure to test over-capitalization and over-punctuation; complexity in the area of sentence structure; and extreme generality and brevity. The other nine test users (20 percent of the total) did not rate it or comment upon it.

III. Teacher Evaluations

Primary I Battery (Form A, Grade 1.5)

In the opinion of the teacher who evaluated it, this is a very good measure of the language skills of early primary grade children. The items in Test 1 (Word Knowledge) provide a simple picture accompanied by four words which might describe "what the picture is about." The children are to indicate their choice of the appropriate word by making a cross (X) in the box

adjacent to it. This teacher felt that the slow or culturally disadvantaged child might have trouble identifying those pictures which are far removed from his immediate environment. Nevertheless, she commended this section and Test 2 (Word Discrimination) for their "excellent basic vocabulary," their measurement of "the ability to think," and their "very good arrangement of items." She answered all items on the questionnaire affirmatively and indicated that she would be interested in using the test in her classroom.

Primary II Battery (Form A, Grade 2)

The teacher who evaluated this section of the MAT seemed to consider it a good measure of word knowledge, word discrimination, and spelling, but of limited value as a test of basic language learnings. She listed as strengths its "good format, precise and adequate instructions, adequate accessory materials, [and] authors' . . . respected professional background." However, she felt that it "does not include language learnings [that] we are trying to develop." She suggested inclusion of tests of "informal standard English, plurals of nouns, [and] patterns of sentences." In addition to her major objection to the lack of material relevant to language, she rated the norms too high for her particular group of students. She stated that she would not use the MAT for testing language in her classroom.

Elementary Battery (Form A, Grades 3-4)

This section of the MAT was judged of limited value as a measure of language skills. The teacher who evaluated it acknowledged that it would be easy to administer and to score, but rated it "too advanced for beginning third grade" pupils. Furthermore, she considered the capitalization and punctuation items confusing to students and suspected that "guess or chance" would play a large part in determining scores in these areas. [In this section the student's attention is directed to a given spot where mechanical corrections may be needed; he is asked to make whatever changes are necessary.] This teacher granted, however, that the usage section would discourage guessing by requiring the student to supply the "correct version" of those usages he considers incorrect.

The teacher's strongest objection concerned the extent to which a high score on such a test could be considered a true indication of his tendency to use "the correct English forms in speaking

and writing." She remarked, for instance, that she "would like to see measurement of a child's ability to express an idea." In conclusion, she stated that she would continue to use the MAT with her students, but implied that she would supplement it with other tests of speaking and writing ability.

Intermediate Battery (Form A, Grades 5-6)

The teacher assigned to this section of the MAT apparently formed a favorable opinion of it. She commended it for staying "within a reasonable range of the grade being tested" and for allowing the child to "take a guess." [She did not specify why she approved its encouragement of guessing, however.] She especially liked Tests 3 and 4 (Spelling and Language), which require students to correct the errors they spot. She responded affirmatively to every item on the questionnaire, objecting only mildly to the test's failure to measure creative writing ability. She mentioned that her students have no trouble recognizing unacceptable usage appearing on such a test, but fail to use acceptable forms in their written work. These objections, however, were not strong enough to prevent her from stating that she would use the MAT in her classroom.

Advanced Battery (Form A, Grades 7-9)

The teacher chosen to consider this battery also expressed reservations about parts of it, but arrived at a generally favorable conclusion. She rated the content "adequate," but suggested that "a judgment test . . . be included" in the usage section. In other words, students should be expected to know which usages are "(1) acceptable anywhere, (2) acceptable in formal writing and speaking, (3) tolerated but not approved, and (4) not acceptable." She considered the test a successful measure of the material covered at the ninth grade level; at the seventh and eighth grade levels, however, she ventured that "much of A, B, and C [Usage, Punctuation and Capitalization, and Kinds of Sentences] would have to be guesswork because the material has not been taught" in her school. She also criticized Test 1 (Word Knowledge), much of which she judged "impractical for many students." Despite these objections, however, she answered a majority of the items affirmatively and stated that she would be willing to use the test in her classroom.

High School Battery (Form BM, Grades 9-12)

The teacher who evaluated this section of the MAT seemed to consider it an excellent measure of English ability at the high school level. She responded affirmatively to every item on the questionnaire and especially commended the "coverage of materials students are being taught" and the inclusion of the Sentence Structure subtest. She preferred it to the test she had used in the past and would be interested in trying it with her students.

IV. Published Reviews

Paul T. Dressel, Director of Institutional Research at Michigan State University, East Lansing, Michigan, confines his remarks to the High School Battery of the MAT. Speaking first of the quality of the items, he wonders why the Spelling Test requires students to spell words correctly, but makes no adjustment for this in scoring. Like several of the teachers who evaluated the MAT, he objects to its apparent encouragement of guessing. Further, he considers the procedure for responding in the Language Test unnecessarily time-consuming. In general, he concludes that the items are "competently done" (p. 57). He notes, however, that "the emphasis is clearly on skills and factual knowledge" and regrets that "items carefully constructed to require critical thinking of all students are not to be found" (p. 57).

Turning to administration and interpretation, Mr. Dressel observes that the instructions for administering and scoring are detailed and clear and that adequate information for interpreting scores is provided. He suggests, however, that more evidence be included to substantiate the author's claim of curricular and content validity. He notes that extensive data are provided on item analyses, test reliabilities, and intercorrelations among subtests. Although he describes reliabilities as "generally adequate," he wonders why separate scores and norms are provided for subtests which appear to measure similar abilities (i.e., Language Study Skills and Social Studies Study Skills).

Next, Mr. Dressel discusses the difficulties encountered by the MAT as a curriculum-oriented battery. First of all, it is almost impossible for a test of this sort to adjust to the "lack of a common sequence of topics in any field of study in the high schools" (p. 58). Furthermore, any curriculum-oriented instrument must necessarily reflect "the traditional curricular emphases of many secondary

schools . . ." (p. 59). He concludes, then, that although the High School Battery of the MAT "fairly adequately test[s] the basic skills and knowledge which [it] undertook to cover," it "cannot be regarded as a significant improvement over" the *Essential High School Content Battery*, which it is designed to replace (p. 59). The MAT might prove useful for guidance or for a general survey of competencies, but Mr. Dressel recommends the ITED or STEP Series to teachers concerned with improvement of instruction or the curriculum.

Henry S. Dyer, Vice-President of the Educational Testing Service, Princton, N. J., would agree with Mr. Dressel that the content of the MAT reflects what the publisher thinks the curriculum is, rather than what it ought to be. He notes that extensive research seems to have been undertaken in preparing the content, especially that of the Word Knowledge and Spelling Tests. However, the content itself suggests "that the schools are still putting a massive emphasis on the rote learning of information and skills, and paying little heed to the development of the more complex cognitive processes normally associated with the maturing mind" (p. 60). In fact, only one-fourth of the items for grades 5-12 "make any demand on the pupil's ability to reason and solve problems" (p. 60). Furthermore, Mr. Dyer, like Mr. Dressel, objects to the method of scoring for the Spelling Test.

Turning to a discussion of the statistical characteristics, Mr. Dyer suggests that more information concerning the degree to which "individual items are contributing to the measurement process" (p. 61) be included. He considers the reliability data adequate, but feels that "the form in which they are reported for the five pre-high school batteries leaves something to be desired" (p. 61). For example, "three kinds of information required for the interpretation of the reliability coefficients" (p. 61) are given for the High School Battery, but are omitted for the others. Furthermore, certain of the Advanced Battery's language tests appear to be unreliable.

Next, Mr. Dyer discusses the national norms. Although he acknowledges that much effort appears to have been devoted to their preparation, he questions the value of norms based upon the participants' willingness to be included. Moreover, he doubts that national norms *per se* are of much value. In his opinion, the most they can provide is "a convenient but arbitrary scale for rendering

scores across tests more or less comparable" (p. 61). He also criticizes the publishers for continuing to perpetuate "the myth that the so-called 'grade equivalent scale' has any normative meaning," since "the very notion of a 'grade' . . . is a glaring example of the fallacy of misplaced concreteness" (p. 61). Another "ancient mistake" made by the publisher is encouragement of the comparison of a student's "achievement" with his capacity (p. 62). Mr. Dyer commends the publishers however, for encouraging test users to produce local stanines and percentile ranks.

In summary, Mr. Dyer seems to feel that although the MAT is carefully prepared, much of its content and statistical information fall short of expectation. His review implies that all but the very conservative users will find the battery outdated.

[Note: *The Sixth Mental Measurements Yearbook* also contains a review of the MAT by Warren G. Findley, Professor of Education and Coordinator of Educational Research at the University of Georgia, Athens, Georgia. He applauds the test's scope, excellent items, careful standardization, and "outstanding" Manual for Interpreting. He feels, however, that the item types used to measure language and spelling need to be improved, and suggests inclusion of tests of effective expression and understanding of language structure. On the whole, he considers the series "superior" (pp. 62-67).]

Objective Test in Grammar

I. General Information

Grades 10-12; 1961; four sections: (I) Parts of Speech, Tense, Person and Number, Grammatical Usage; (II) Grammatical Correctness, Sentence Recognition; (III) Agreement; (IV) Diction, Punctuation; separate answer sheet required; scoring key provided; no manual; no instructions for administering; no data on reliability and validity; norms provided on request; Nellie F. Falk; The Perfection Form Company.

II. Use in Wisconsin

The OTG is used by five, or 2 percent, of the respondents to the first questionnaire who use standardized tests. One of these, or 20 percent, rated it satisfactory but gave no reason why; two, or 40 percent, judged it unsatisfactory because it is too long and provides "questionable" results; two, or 40 percent, did not rate it.

III. Teacher Evaluation

The teacher who evaluated the OTG was hindered by a lack of instructions for administering and for interpreting results. [The editor requested this information from the publisher, but received only a brief description of the norms.] Thus, this teacher's judgment is based solely upon content. He regarded the sections dealing with agreement and diction as the test's "strong areas." In his opinion, "the test's chief weakness lies in its requiring the student to apply labels in 66 of the 150 items . . . ." He also questioned the validity of several of the items, not specifying which, and criticized the format, which requires the student to "turn back to the preceding page to refer to an answer symbol key . . . ." Further, he felt that "too little emphasis" is placed on "sentence sense, syntax and structure."

Since this teacher received no information concerning the abilities the OTG purports to measure, he could not answer item 2 on the questionnaire. On the basis of content, however, he responded negatively to the other items. He did not consider "items requiring the student to label the parts of speech" a valid part of

the English curriculum, nor did he feel that the content is related to "informal standard English." He judged the test long enough, but not comprehensive enough, to be reliable, and felt that it should provide a "means to evaluate achievement in closely related areas of language study." In short, he formed an unfavorable opinion of the OTG, which he stated he would not use with his students.

IV. Published Reviews

No published reviews of the OTG were found.

The Purdue High School English Test

I. General Information

Grades 9-12; 1931-62; modified from the *New Purdue Placement Test in English*; 6 scores: grammar, punctuation, effective expression, vocabulary, spelling, total; IBM and MRC; Forms 1, 2 ('62); manual; separate answer sheets required; $4.20 per 35 tests; 36 (45) minutes; H. H. Remmers, R. D. Franklin, G. S. Wykoff, and J. H. McKee; Houghton Mifflin Company.

The stated purpose of the PHET is to sample the knowledge of "good English" possessed by high school students and college freshmen. Norms based on both part scores and total scores are provided; the latter are listed according to sex and grade. Items were selected from the *New Purdue Placement Test in English* after administration to 370 students in grades 9-12; each item was then analyzed in terms of its difficulty and its "discrimination index" (p. 21). Claims of validity are based upon correlation of total scores to self-reported grades. Reliability data are computed from scores of "a systematic sample of 400 students in the norm group." Standardization is based upon a representative sample of 2,000 for the spring norms and 1,000 for the fall norms. Sex, grade, region, and type of residence were taken into account. Norms for college freshmen are based upon administration of both forms to the 2,200 freshmen enrolled in Freshman English at Purdue University in September, 1962.

II. Use in Wisconsin

The PHET is used by eight, or 3 percent, of those respondents to the first questionnaire who employ standardized tests. Seven, or 87.5 percent, of these rated it satisfactory because it is helpful in diagnosis of basic strengths and weaknesses, comprehensive, and useful in student placement. None judged it unsatisfactory, but one did not rate the test or comment on it.

III. Teacher Evaluation

The teacher who evaluated the PHET considered it a "good test." She answered all questionnaire items affirmatively, commenting only that juniors and sophomores might not perform well on the

test. She would attribute this fact to inadequate preparation, not to a fault in the test. She felt that her seniors would do "fairly well." She had not used this test at the time of writing, but stated that she would consider adopting it.

IV. Published Reviews

Charlotte Croon Davis of the Test Research Service, Bronxville, New York, considers the PHET inferior in subject matter coverage, item content, and techniques of testing. She describes the selection of items from the *New Purdue Placement Test in English* [see Part I above] and attributes the inadequacy of items to the weight placed upon statistical evidence as the basis of selection.

The reviewer continues with a discussion of specific instances of content inadequacy appearing in the Effectiveness of Expression subtest. Twenty sentences are to be marked *r* or *w* depending on whether or not they are "clear and effective" (p. 575). Those to be marked *w* "are seriously faulty in structure and consequently test . . . at a rather low level," while some to be marked *r* are "not wrong, not unclear, but not particularly effective" (p. 575). To avoid confusing the able student, the reviewer suggests abolishment of the *r-w* dichotomy (which also encourages guessing). Instead, she suggests that students might be presented with two or more ways of expressing an idea and asked to choose the more (or the most) effective (p. 575). The *r-w* method is also used in Parts 1 and 2 (Grammar and Punctuation), which might be more effective if students were required to select the correct responses from a group of two or more.

Miss Davis does commend some features of the test; e.g., the clear format, convenient time period, self-marking answer sheet, and nationally representative norms sample. However, her strong objections to content prevent her recommendation of the PHET to teachers of English.

Benjamin Rosner, Director of the Test Development Division of the Educational Testing Service, Princeton, N. J., notes in his review that the normative emphasis implies that total test performance is "the more significant index" (p. 576) of achievement. While he considers the reliability of the total score acceptable, he is less confident about the reliabilities of the subtest scores, especially those for Parts 1, 2, and 3. He also suggests inclusion of

proof of content validity and evidence supporting the claim that this is a "power test" which all students should be able to complete in the allotted time. He attributes the inadequate reliabilities of subtest scores to two factors: minimum length and the two-choice (*r-w*) item format. For these reasons, he recommends that the test user ignore the norms for the subtest scores and discontinue use of subtest scores for diagnostic purposes. Until reliabilities are increased and content validity improved, he suggests that users consider only the total score, which "yields an acceptable measure of . . . the mechanics of English usage confounded by a rather inadequate assessment of effective English expression" (p. 576).

(Cooperative) School and College Ability Tests

I. General Information

Grades 4-6, 6-8, 8-10, 10-12, (12-14, 15-16); 1955-63; 2 scores pertinent to language arts: verbal, total; IBM, NCS, and Grade-O-Mat; SCAT-STEP student report, profile, supplements (1958, 1962, 1963); SCAT directions for administering, manual for interpreting, technical report; separate answer sheets required; $4.00 per 20 tests; see publisher's catalog for other prices; 70 (95) minutes; Cooperative Test Division.

School Ability Test, 1955-63, 2 forms, contains 4 levels:

1. Level 5. Grades 4-6; Forms 5A, 5B, ('56).
2. Level 4. Grades 6-8; Forms 4A, 4B, ('56).
3. Level 3. Grades 8-10; Forms 3A, 3B, ('56).
4. Level 2. Grades 10-12; Forms 2A, 2B, ('55).

College Ability Test, 1955-63, contains 2 levels (grades 12-14; 15-16). [Note: There are 71 references to articles pertaining to the SCAT series included in *The Sixth Mental Measurements Yearbook*, pp. 715-16.]

II. Use in Wisconsin

According to the results of the first questionnaire, 24, or 8 percent, of the respondents who use standardized English tests employ the SCAT series. Ten, or 42 percent, rate it satisfactory for various reasons: ease of administration; applicability of norms; and accurate measurement of the ability to deal with mechanics and of verbal ability on the basis of vocabulary. Two, or 8 percent, consider the test unsatisfactory because it does not help identify individual weaknesses in specific areas and because the norms sample is too small. Since 12, or 50 percent, of the test users did not rate it, however, the significance of the above figures is questionable.

III. Teacher Evaluations

Level 5 (Form 5B, Grades 4-6)

The teacher evaluating this section of the SCAT applauded its helpful directions, clear print type, and "interesting" verbal material. She accepted the content as it stands, but did not consider the test long or comprehensive enough to provide a reliable measurement of achievement. In her words, "it appears to try to cover too broad a base without adequate sampling." Because the norms are "geared to school systems with a larger population than ours," she judged these norms unrealistic as a measure of her students' performance. In addition, she did not consider the supplementary materials sufficiently clear. For these reasons, she stated that she would not use the test with her students.

Level 4 (Form 4B, Grades 6-8)

The evaluator of this portion of the SCAT listed four major strengths: the convenient time element; "clear and concise" directions; extensive coverage of vocabulary; and adequate materials for evaluation. She considered the content important to the English curriculum, but criticized its narrow scope. In her words, "many areas . . . which in my estimation are just as important" are not touched. For instance, "usage skills, . . . comprehension, and understanding" are not included. Because most of her students are "emotionally disturbed under-achievers," this teacher would use a different test in her classroom. She suggested, however, that the SCAT might be used to advantage for placing students into small groups within the classroom.

Level 3 (Form 3B, Grades 8-10)

No report.

Level 2 (Form 2B, Grades 10-12)

The teacher who evaluated this level of the SCAT judged it satisfactory in all respects. She especially liked the convenient time limit, the generality of the content, and the variety of test forms. She considered the content a valid part of the English curriculum, "unless one is interested in testing for some special skills other than those indicated . . . ." The SCAT is used in her school, where the national norms are found valuable in comparing students going

on to college with students from other parts of the country. In summary, both she and her school seem to accept the SCAT as an adequate measure of verbal ability.

IV. Published Reviews

Russel F. Green, Associate Professor of Psychology at the University of Rochester, Rochester, New York, considers the SCAT a very good set of scholastic aptitude tests. However, he raises some objections which may be of interest to test users. For instance, he wishes that the publishers had been more specific in their statements of purpose and their descriptions of what is measured. In their apparent reluctance to discuss the specific abilities tested by an instrument unrelated to course content, the publishers have failed to defend the total score as a predictor of academic performance. In addition, while validities are generally well demonstrated, the test appears to predict achievement better among women than among men. The inclusion of the verbal score is questionable, since English grades seem to be predicted more accurately by the total score. Mr. Green also notes that no compensation for probable "speededness" is made in computing the percentile bands. Finally, he suggests that the test for level 5 (grades 4-6) not be used for discriminating strong and weak fourth graders, for whom it appears too difficult. These objections, however, do not prevent Mr. Green from endorsing the SCAT series as "a model of how such a series should be planned, developed, standardized, and validated" (p. 453). He strongly recommends it as a measure of scholastic aptitude and predictor of academic performance.

SRA Achievement Series: Language Arts

I. General Information

Grades 1-2, 2-4, 4-6, 6-9; 1954-64; Subtest of *SRA Achievement Series;* 2 editions; battery teacher's handbook for both editions; Louis P. Thorpe, D. Welty Lefever, and Robert A. Naslund; Science Research Associates, Inc.

Forms A and B. Grades 2-4, 4-6, 6-9; 3 scores: capitalization-punctuation, grammatical usage, spelling; IBM for grades 4-9; 3 levels; administrator's manual; technical supplement; pupil progress and profile charts; separate answer sheets required in grades 4-9.

1. Grades 1-2. Form A ('58); examiner's manual; $3.50 per 20 tests.
2. Grades 2-4. Forms A ('55), B ('57); examiner's manual; $2.00 per 20 tests; 70 (95) minutes in 2 sessions.
3. Grades 4-6. IBM; Forms A ('54), B ('56); examiner's manual; $2.15 per 20 tests; 75 (90) minutes.
4. Grades 6-9. IBM; Forms A ('55), B ('56); examiner's manual; $2.00 per 20 tests; 60 (75) minutes.

Forms C and D. Grades 2-4; 4 scores: same as Forms A and B plus total; Forms C ('55 revised '63), D ('57 revised '63); tests essentially same as Forms A and B except for format; examiner's manual for each form; test coordinator's manual; pupil progress and profile charts; $2.00 per 20 tests; 60 (85) minutes in 2 sessions.

The publishers claim in their teacher's manual that this battery of tests forms an integrated program for measuring the educational development of students in grades 1 through 9 in the basic areas of the curriculum. They suggest that the ITED be used in grades 9-12 to provide "a continuous program of measurement" throughout the grades. Three main purposes of the SRA series are stated in the manual:

1. "To enable teachers and counselors to keep intimately and reliably informed of the educational development of each student.

2. "To provide an objective and comprehensive description of the educational development of groups of students.
3. "To provide a means for curriculum evaluation and planning."

The content of each part of this battery is based upon a careful study of the literature and instructional materials in the basic curricular areas. The publishers state in the manual that the Language Arts Test is geared to measure the student's actual use of the English language instead of his ability to memorize rules or definitions.

II. Use in Wisconsin

Fifteen, or 5 percent, of those users of standardized tests responding to the first questionnaire employ the *SRA Achievement Series*. Six of these, or 40 percent, rate it satisfactory for the following reasons: it is "up-to-date"; it measures thinking ability; it helps students organize and write for a given purpose; it is a good test of mechanics and usage. Three test users (20 percent of the total) rated it unsatisfactory but gave no reasons for their judgment; six (40 percent) neither rated it nor commented upon it.

III. Teacher Evaluations

Grades 1-2 (Form D)

The evaluator of this section of the SRA test listed three chief strengths: the items testing visual and auditory discrimination of initial and terminal sounds of words; the vocabulary test; and the continuity achieved by following this test with the ITED at the high school level. This evaluator considered the failure to describe in detail the standardization sample a major weakness. The manual states that 71,199 students in 252 schools located throughout the United States comprise the sample, but gives no information about the specific cities included or the size of the school systems involved.

This teacher answered questions 2-4 affirmatively. However, he and his committee considered some parts of the test, especially the reading section, "too difficult." They questioned the validity of the norms because of the "vague description of the sampling procedure," judged the manuals "rather unwieldy," and did not recommend the test for use in their school.

Grades 2-4 (Form D)

Two teachers evaluated this section of the SRA test. The first responded affirmatively to all questions, listing as chief strengths the "excellent size of print" and the thorough coverage of material which is actually taught. This teacher suspected that the test might be too difficult for the below average student, but stated, nevertheless, that she would use it with her students.

The second teacher considered the test easy to administer, but objected to its "drab" appearance and "lack of color." In contrast to the previous evaluators, he wondered "if the instrument is too easy" and suggested that it should include "the use of indexes, tables of contents, charts, alphabetical lists, [and] dictionary usage." He acknowledged that "it does give one a fairly good idea of what level a child is working at," but felt that a classroom teacher could "do just as well using his own judgment."

Grades 4-9 (Form D)

This portion of the SRA series was rated highly by the teacher who evaluated it. She praised the "clarity of directions," "interesting form and content," arrangement and presentation of items, item analysis report, and multi-level concept. Her objections were minor. She found it difficult to keep her eye on the correct row while taking the spelling test and suggested that students be given "a marking device to slide down the row" as they work; she suggested that students be told before taking the test that reading ahead or back "to make sure of the proper choice" is permitted and will not prevent completion of the work. This teacher responded to each item of the questionnaire with a strong affirmative and stated that this was the first test she had examined that attempts to test actual language rather than memorization and definition. She noted that "it does not claim to measure creativity," but she expressed doubts that creativity can be measured by this kind of test. This examiner rated the SRA series superior to the ITBS, which she now uses, and felt that it would be excellent for use with her fifth grade class.

IV. Published Reviews

Miriam M. Bryan, Associate Director of Test Development for the Educational Testing Service, Princeton, N. J., begins her review of Forms A and B of the *SRA Language Arts Test* by de-

scribing a major modification made in the 1964 edition. A recall-type spelling test has been added to the grades 2-4 battery to permit plotting of the growth of spelling achievement for all grades in which spelling is taught. The reviewer considers the spelling words "sensibly chosen," although she finds the test extremely difficult for second graders and "of middle difficulty" (p. 578) in the first semester of the fourth grade. She also questions aspects of the spelling tests for grades 4-9, approving the presentation of items in context but suggesting the inclusion of more items in a shorter context. Miss Bryan advocates, too, a different arrangement of responses in the multiple choice items: placement of the word tested in the same position on the answer sheet in which it appears in context. This, she contends, would be less confusing to students than the present placement in varying positions.

Turning to the capitalization, punctuation, and grammatical usage sections, Miss Bryan judges coverage of these areas "quite complete," although parts of the grades 2-4 battery again appear "somewhat sophisticated" (p. 578) for this age group. In the same battery, the lack of precision in the underlining of items creates some confusion between possible responses. Furthermore, at all levels, there is a considerable amount of inconsistency between the punctuation required in a particular item situation and punctuation used elsewhere in the test" (p. 579). This involves the use of commas to punctuate nonrestrictive adjective clauses and to set off introductory adverbial clauses, even at the primary level. Miss Bryan suggests modification of these inconsistencies and questions the inclusion of items concerning the use of a comma before "and" in a series, a matter about which language experts do not agree.

Miss Bryan finds the accessory materials "complete and convenient" (p. 579) and praises the care with which the test was prepared and standardized. She feels, however, that instructions for proper placement and manipulation of the multi-level answer sheets might be included. She considers her criticisms minor in view of the test's generally high quality, and she ranks it "high among existing tests in language arts for the grade levels for which they are designed" (p. 579).

SRA High School Placement Test

I. General Information

Entering 9th grade students; 1957-63; 1 score pertinent to language arts: language arts achievement; new form issued annually; 3 tests in use in 1963; Series 64K ('62), Series 63K ('61), Series 63A ('60); optional Catholic religion test available: examiner's manuals for all series; technical reports for all series; profile leaflet; separate "Docu Tran" answer sheets required; tests loaned only; examination fee: $1.10 per student (includes scoring service, reporting of normed scores and local norms); total battery requires 185 (230) minutes for 64K, 175 (215) minutes for 63K and 63A; Science Research Associates, Inc.

II. Use in Wisconsin

The *SRA High School Placement Test* is used by seven, or 2 percent, of the respondents to the first questionnaire. Four, or 57 percent, rated it satisfactory because student scores correlate with other test scores and with actual performance, and because the battery indicates student needs in the high school English program. One user found it helpful in establishing a remedial reading program for students who had not attained an eighth grade reading level. The test was judged unsatisfactory by two users, or 29 percent, of the total; one user did not rate it. No reasons were given for either rating.

III. Teacher Evaluations

No report.

IV. Published Reviews

Walter N. Durost, Associate Professor of Education at the University of New Hampshire, Durham, New Hampshire, notes that Series 64K of the SRA-HPT is intended for use in parochial schools for determining student acceptability, placing students, and evaluating achievement. He wonders why the publishers have paid no attention to the test's use in public schools, and suggests that they do so. Although he approves the selection of subtests, Mr.

Durost considers the language arts test of "questionable" value and feels that item quality "leaves much to be desired" (p. 89). For example, the word-reasoning test overemphasizes nouns, lacks clarity as to how words function in context, and employs imprecise synonyms whose meanings only approximate those of the stimulus words. In all parts of the test, in order to score well the student must respond as the authors expect, rather than as he thinks correct.

Although Mr. Durost considers the examiner's manual "well organized and reasonably clear and explicit" (p. 90), he suggests that the test be administered in several sittings. [Total testing time is three hours, fifty minutes.]

Turning to standardization, Mr. Durost recommends that the test be normed by "administration to large groups of parochial school pupils" (p. 91) rather than by the present method of equating it with the *SRA Achievement Series*. He believes that "a serious technical error" was made by equating the educational ability score to the *Otis Quick-Scoring Mental Ability Tests* to obtain IQ equivalents. In the first place, the Otis IQ's are not derived by the method the publishers suggest. Furthermore, the sample used to obtain the IQ norms seems to represent several grades and ages, rather than grade 9 only. Most seriously, the mental ages in the Otis test were not derived for the purpose of computing IQ's.

Mr. Durost concludes by apologizing for the negative tone of his review by noting that no objective test ever measures up to the ideal standard in the reviewer's mind. The *SRA High School Placement Test*, he states, "is not a bad test as such tests go" (p. 92).

Charles O. Neidt, Professor of Psychology at Colorado State University, Fort Collins, Colorado, considers Series 63A, 63K, and 64K (*The Sixth Mental Measurements Yearbook* reads "64A," but the reviewer twice refers to "64K" and has been assumed correct) of the SRA-HPT "a satisfactory measure of general scholastic aptitude" (p. 92). Like Mr. Durost, he describes the procedure followed in expressing educational ability scores as IQ's, and suggests that educational ability raw scores be converted instead to derived IQ's. Both reviewers, then, question the effectiveness of the procedure now used, although their suggestions for improvement differ.

Mr. Neidt finds the items generally "well-constructed" (p. 93), but regrets that no item statistics are provided. He notes that correlations of scores with course marks are generally high, but recommends a new standardization for the test. Like Mr. Durost, he suggests that the norms sample include parochial school students. Noticing that the mean scores of girls tend to be higher than those of boys, he recommends careful inspection of item statistics or preparation of norms according to sex to compensate for this difference.

Mr. Neidt considers the greatest shortcoming of the SRA-HPT to be its lack of a measurement of science achievement. Nevertheless, he believes that the three present editions represent significant improvements over earlier editions.

Sequential Tests of Educational Progress

I. General Information

Grades 4-6, 7-9, 10-12, 13-14; 1956-63; IBM and Grade-O-Mat; 2 tests pertinent to language arts: writing, essay; Forms A, B ('57) of writing test; Forms A, B, C, D of essay test; directions for writing test; examiner's handbooks for essay test; interpretive manual for writing test; technical report; 1958, 1962, 1963 SCAT-STEP supplements; teacher's guide; SCAT-STEP profile and student report; no data on reliability for Form B; separate answer sheets required for writing test; $40.00 per 20 tests (except essay test); $1.00 per 20 essay tests; see publisher's catalog for other prices; 35 (40) minutes for essay test; 70 (90-100) minutes for writing test; Cooperative Test Division.

The interpretive manual accompanying the STEP series states that it is designed to measure "the broad outcomes of general education, rather than the relatively narrow results of any specific subject-matter course." The focus is upon "solving new problems on the basis of information learned," with provision for continuous measurement of the development of individual students.

The Writing Test includes items which fall into the five categories of organization, conventions, critical thinking, effectiveness, and appropriateness. Students are required not only to recognize errors but to select appropriate revisions. Passages "are drawn largely from materials actually written by students in schools or colleges — assignments which, by and large, were graded poor or failing." The tests cover four levels of difficulty; each contains 60 multiple-choice items. No grade designations appear on the booklets, and administration of different levels to students of differing abilities is encouraged. Instructions for administering are the same at all levels so that different levels may be given simultaneously. The manual describes the various uses of individual and group results.

The Handbook for Essay Tests describes them as "free-response tests" of writing ability. The student is presented with a brief paragraph setting forth a topic; he is given 35 minutes to read the paragraph and to plan and execute his response. His

writing performance is judged by comparing his paper with previously rated student papers. As with the writing series, tests at four levels of difficulty are available. In the Essay series, however, four alternate tests are provided for each level so that a student may be tested at the same level more than once. Topics are appropriate to specific educational levels; teachers are cautioned that students may regard topics designed for levels higher or lower than their own as unsuitable.

II. Use in Wisconsin

The STEP series, the second most widely used in Wisconsin, is employed by 103, or 34 percent, of those respondents who use standardized tests. Sixty, or 58 percent, of these rate it satisfactory for the following reasons in order of frequency:

1. Eleven consider the availability of state norms an asset.
2. Six judge the test an adequate predictor of language arts ability.
3. Two use the test as a general guide for programming.
4. Others cite its correlation with curricular material, comprehensive scoring data, ease of administration, measurement of critical thinking, and focus upon English as part of a student's total education.

One teacher commented that the STEP "seems to subscribe to a more liberal view of a changing language." Another commended the Essay Test as "a measure of writing ability as opposed to recall of rules of grammar and usage." This teacher also noted that the STEP furnishes teachers with a starting point for individual teaching. In addition, a committee at Racine, Wisconsin, investigating standardized tests found the STEP series "more promising and less time-consuming" than others they considered.

Only six STEP users, or 6 per cent, of the total, rated the series unsatisfactory. Two found it inadequate for diagnosing specific weaknesses. Another judged it "dogmatic concerning mechanics." Still another considered the norms sample (5,000) too small and felt that "something should be done to rectify the situation." Thirty-seven, or 36 percent, of those who use the test did not rate it.

III. Teacher Evaluations

Forms 4A and 4B (Grades 4-6)

This level of the STEP Essay and Writing Tests was rated favorably by the teacher who evaluated it. She felt that it would provide "a reliable indication of achievement" and would be helpful in determining "where a student stands in his section or class." She answered all questions affirmatively, commenting only that "the examiner would have to study carefully the information on score tabulations and correlations to be able to use [the tests] to the best advantage." She doubted that "the ordinary classroom teacher" would have time to achieve a thorough understanding of her class's national standing. However, she stated that she would consider adopting the STEP series in her classroom.

Forms 3D and 3B (Grades 7-9)

This level of the STEP series was rated favorably by the teacher who examined it, although he expressed certain reservations. He judged the Essay Test "excellent," especially in its use of types of writing appropriate to the junior high school level. The Writing Test was considered "good in that it requires application of principles," but " 'overweighted' with errors in an almost negative tone." Moreover, it "requires distinctions about bias on levels not accomplished by seventh graders." Although the teacher answered questions 2-4 and 6-7 affirmatively, he did not consider the tests comprehensive enough to provide a reliable indication of aptitude and achievement. He commended certain features of the supplementary materials, such as the oral directions and the five-minute "thinking" period before writing in the Essay Test. He wondered, however, whether students might be confused by the contradictory instructions to: (1) answer all questions, and (2) use extra time to restudy and answer doubtful items. At the time of writing this teacher did not use the STEP series in his classroom. Although his objections to the Writing Test would prevent his adoption of that section, he stated that he would like to try the Essay Test in his ninth grade class.

Forms 2D and 2B (Grades 10-12)

Two teachers evaluated this level of the STEP series. The first, like the teacher who rated Forms 3D and 3B, judged the

Essay Test "excellent" but expressed reservations about the Writing Test. She especially approved the essay topic, which "certainly tests a student's ability to look beneath the speech and actions of an individual and to see him as he really is. It also tests his ability to organize and to bring his ideas forward into a concluding statement." She described the material covered by the Essay Test as "analytical writing, depth writing, specific details, [and] conclusions supported by evidence." She felt that it covers "a good variety of writing skills," but should also include parallel structure, run-on sentences, and the use of figurative speech. She considered it "free of the too formal, rather stilted language" found in other objective tests, but suggested that "a few of the articles might have . . . a more mature writing type of analysis." Despite her objections, however, she answered all questions with general affirmatives and stated that she would use both tests in her classroom.

The second teacher to evaluate this level of the STEP series considered the Writing Test "as successful as any objective test which tries to measure writing can be." Although she doubted that tests of this kind "can actually measure ability to organize materials or to write effectively," she granted that the STEP Writing Test "*does* make the student think" and "analyze [his] answers carefully." She objected that "there was too fine a point between right and wrong" in some of the items, but felt that the Writing and Essay Tests together would "provide a realistic indication of student aptitude and achievement." In general, she approved the series, which she stated she "would like to use" with her students.

IV. Published Reviews

Harold Seashore, Director of the Test Division, The Psychological Corporation, New York, New York, considers the format of the booklets, the "universal" answer sheets, and the general flexibility of the STEP series "strong feature[s] of the test battery" (p. 101). A possible exception, however, is the Essay Test, which requires students to write in the booklets. He notes that each level of the Essay Test requires a separate manual, and suggests that these be reduced from the present 144 pages to 64 pages by combining their identical content. The same could be done for the *Manual for Interpreting*. Another suggested change is inclusion in the booklet, *Directions for Administering*, of sample items more closely resembling those in the Writing Test. In general, Mr. Seashore seems to find the manuals a valuable source of "functional

information," but regards as their chief shortcoming that they "overpower one with redundancy" (p. 102). He suggests that this situation can and should be speedily remedied.

Next, Mr. Seashore notes that previous criticisms of the STEP series, pertaining to the adequacy of the manuals and to the system of converted scores, seem to have gone unheeded. Addressing the publishers, he asks why new data on reliability, validity, intercorrelations between subtests below the college level, and the relation of STEP to other tests have not been included. He criticizes the use of "situational" items designed to "simulate real life problems" (p. 103). Such items, in his opinion, rely too heavily upon reading ability and might even serve as a "crutch" to students, who would "reflect a higher order of . . . understanding" by sensing the nature of a problem without reference to the "situation" (p. 104).

Turning to the standardization program, Mr. Seashore rates the documentation of procedure "excellent," but criticizes the limited size of the norms sample and the publishers' apparent "tendency to be satisfied with statistical manipulations" (p. 104). He suggests that large samples of real cases rather than "grossly estimated" (p. 105) values should be used in presenting statistical data.

In conclusion, the reviewer questions the general value of non-curriculum-oriented tests such as STEP. Users of the series should realize that it will not evaluate teacher effectiveness, curricular adequacy, and individual student growth; other instruments should be substituted if these purposes are intended. Once its limitations have been acknowledged, however, the STEP Series can be a serviceable device.

Hillel Black, Senior Editor of *The Saturday Evening Post*, New York, New York, states that the STEP Writing Test performs "a grave disservice to the teaching of English composition" (p. 592). Of the five skills the test claims to measure, only "conventions" is actually measured, and this is done successfully only "when the mental process is an act of memory involving such mechanical tasks as spelling and punctuation" (p. 592). In the reviewer's opinion, the "organization" section fails to measure the ability to create ideas and to select and order facts. Instead, the student is asked to take "facts and ideas already organized for him and then [to perform] what may be called minor editing, such

as rearranging or deleting sentences" (p. 593). The ability to make such editorial revisions, argues Dr. Black, should not be equated with the ability to write well. According to Mr. Black, the test totally prevents students "from offering any original concepts composed in an original manner" (p. 595). The best test of writing skills, suggests Mr. Black, is not an instrument such as this, but writing itself.

Dean A. Allen, writing in the *Personnel and Guidance Journal*, Vol. 42, pp. 298-303, November, 1963, agrees with Mr. Black that the student-composed passages in the STEP Writing Test "are almost unvarying in their poor quality and trivial content" (p. 595). Not only is the choice of topics disappointing, but the passages are full of errors which are not singled out for revision. This makes taking the test an ordeal for the student and "all too often makes [his] choice of best answer hinge on the relative importance he assigns to consistency of style vs. good English" (p. 595). Careful editing is needed, moreover, to correct those items which contain no acceptable good answer, two or more good answers, or a wrong answer keyed as correct.

Mr. Allen notes further that correlations between the Writing Test and other STEP and SCAT tests "confirm the impression that . . . [it] may be measuring general scholastic aptitude rather than writing skills as such" (p. 595). He suggests that correlations between alternate forms and more information on reliability and validity should be provided. Noting that the publishers make all comparisons on the basis of percentiles, which can be obtained easily from raw scores, he questions the value of converted scores and implies that these might be eliminated.

In conclusion, Mr. Allen praises the careful planning and detailed presentation of technical material that characterizes the STEP Writing Test, but suggests, as does Mr. Black, that it falls short of actual composition as a test of writing ability.

Stanford Achievement Test (1964 Revision)

I. General Information

Grades 1.5-2.5, 2.5-3.9, 4-5.5, 5.5-6.9, 7-9; 1923-64; 1953 revision still available; subtests in spelling and language for grades 4-9 available as separates; IBM and MRC for grades 4-9; reliability data for one (unspecified) form only; separate answer sheets may be used for grades 4-9; Truman L. Kelley, Richard Madden, Eric F. Gardner, and Herbert C. Rudman; Harcourt, Brace & World, Inc.

a) Primary I Battery. Grades 1.2-2.5; 2 scores pertinent to language arts: vocabulary, spelling; Forms W, X ('64); manual; $5.65 per 35 tests; entire battery requires 127-160 minutes in 5 sessions.

b) Primary II Battery. Grades 2.5-3.9; 3 pertinent scores: word meaning, spelling, language; Forms W, X ('64); manual; $5.80 per 35 tests; entire battery requires 185-235 minutes in 7 sessions.

c) Intermediate I Battery. Grades 4-5.5; same scores as in *b*; manual; supplementary directions for use with IBM, MRC answer sheets; $8.25 per 35 tests of partial battery (Form W); entire partial battery requires 201-230 minutes in 5 sessions.

d) Intermediate II Battery. Grades 5.5-6.9; same scores as for *b* and *c*; manual; supplementary directions for use with IBM, MRC answer sheets; prices for partial battery (Form W) same as for *c*; entire partial battery requires 192-219 minutes in 5 sessions.

e) Advanced Battery. Grades 7-9; scores same as for *b*, *c*, *d* with the omission of Word Meaning; manual; supplementary directions for use with IBM, MRC answer sheets; prices same as for *c*, *d*; entire partial battery (Form W) requires 178-201 minutes in 4 sessions.

The manual entitled *Directions for Administering* states that the SAT was "developed to measure the important knowledges, skills, and understandings commonly accepted as desirable outcomes of the major branches of the elementary curriculum."

Scores are comparable from subject to subject and from grade to grade; the series is designed to be used for "improvement of instruction, pupil guidance, and evaluation of progress." It is emphasized that "persons with little or no training in the use of standard tests" will find the tests easy to administer, score, and interpret.

The Word-Meaning Test measures "knowledge of synonyms, . . . simple definitions, . . . ready associations, . . . [and] higher-level comprehension of the concepts represented by words . . . ." Words frequently used and encountered by students were selected and appropriateness "was checked by reference to the available word counts." Spelling is tested using the multiple-choice format: the student must choose from four words the one spelled incorrectly. The publishers claim that results of this type of test "correlate to a very high degree with results of dictation-type tests." The likelihood that brief exposure . . . to misspellings . . . will have any tendency to fix the incorrect spelling of any one of them in the pupils' minds" is discounted. The bases of word selection were "several research studies . . . , the work of graduate students in listing new words found in magazines, newspapers, and in children's writing . . . , and textbooks in spelling." The Language Test contains exercises in Usage, Punctuation, Capitalization, Dictionary Skills, and Sentence Sense. Items in the Punctuation, Capitalization, and Sentence Sense sections "are presented in a connected discourse, [which] adds interest and provides a more natural testing situation than is achieved with isolated sentences." Scores on the usage part are claimed to be "very useful for group diagnosis." Because "modern usage is occasionally at variance, items on matters that are very controversial have been avoided." The publishers acknowledge, however, that in the present time of transition some controversy is inevitable. They admit that an objective language test may be "somewhat artificial," but claim that this one "affords an adequate appraisal of mastery of those aspects of language which [it] purports to cover."

The SAT was standardized by administration to more than 850,000 pupils in all 50 states. National norms in terms of grade equivalents, percentiles, and stanines are available. Extensive directions for administration, scoring, and interpreting scores are provided, as are statements of reliability and validity and suggestions for using test scores.

II. Use in Wisconsin

Thirty-two, or 10 percent, of those respondents to the first questionnaire who use standard tests employ the SAT. It was rated satisfactory by 18, or 56 per cent, for the following reasons:

1. Four found it "comprehensive."
2. Three praised it for "measuring what is actually taught."
3. Others liked its accurate diagnosis of specific strengths and weaknesses, sequential nature, good standardization program, and accurate measurement of "the ability to deal with mechanics." The same Racine committee that approved the STEP series recommended the SAT for use in the elementary testing program.

Six test users, or 19 percent, considered the SAT unsatisfactory for a variety of reasons. It was judged too brief and too easy, limited in its coverage of skills, and poor at isolating individual problems. [Note: the publishers caution against use of part scores for such diagnosis.] Others considered the emphasis of the Usage section on commonly misused items too strong and results for average students "negligible." Eight, or 25 percent, of its users did not rate the test or comment upon it.

III. Teacher Evaluations

Primary I Battery (Form W, Grades 1.2-2.5)

This section of the SAT was rated favorably by the teacher who evaluated it, although she did mention certain deficiencies. Her evaluation is based upon classroom use of the test as well as upon her responses to the questionnaire. She listed as chief strengths the high reliability and validity figures, the correlation of scores with classroom performance, the "ample" time limits, the availability of comparative forms, the size of the norms sample, and the "information provided for grouping procedures." She considered the test's chief weakness the difficulty of determining "specific difficulties in the areas of reading, spelling, vocabulary and word study skills . . . ." She noted, however, that "at no point have the makers of the test indicated that it is to be used as a diagnostic measure." She judged the test successful in measuring the abilities it claims to measure, but stated that "many of the areas which we emphasize have not been included . . . ." Among these are punctuation, capitalization, and page arrangement. Al-

though "the test actually leans toward a more formal structure," it maintains, in her opinion, a direct relationship to "informal standard English." She considered length and comprehensiveness adequate, but noted that "the number of items could not be lengthened" since "some children show signs of tiring before test sections are completed." Because "a large percentage of [her] children . . . come from educationally deprived homes," she found the stated norms high for her area. Despite these objections, however, she seemed well enough satisfied with the test to continue using it in her classroom.

Primary II Battery (Grades 2.5-3.9)

No report.

Intermediate I Battery (Form Y, Grades 4-5.5)

This portion of the SAT was approved by the teacher who evaluated it. She listed as strengths the large number of items included in each subtest and the size of the norms sample. She felt, however, that "some directions could be clearer" and that "phonics will be difficult for those who were not exposed to it." Her only other objection concerned the extent to which an objective test measures actual writing ability—a skill which the SAT does not claim to measure. With these exceptions, her responses to the questions were generally affirmative. She stated that she would use the test in her classroom "if it were provided."

Intermediate II Battery (Form Y, Grades 5.5-6.9)

The teacher who reviewed this section of the SAT also responded affirmatively to all questions. She liked the format of the test and judged the requirement that students supply the correct word in a sentence in the vocabulary test "conducive to problem solving and writing." Although she would prefer a dictation-type spelling test to the multiple-choice variety and would rate the punctuation and dictionary skills tests too difficult for some fifth graders, she considered the language test "good" on the whole. She stated also that she would like to adopt the test for use in her classroom.

Advanced Battery (Form K, Grades 7-9; this evaluation is based upon administration of the 1953 Revision.)

This section of the SAT was described by the teacher who evaluated it as "a challenge as it reached beyond the levels of some students." She noted that students found the paragraph meaning subtest "unusually difficult." She suggested, however, that the content, especially in the "grammar section," could be more comprehensive and that key sheets and answer sheets should be made to line up more accurately. [Note: these objections may have been met by the 1964 Revision.] All other questions were answered affirmatively. The teacher mentioned that she used the test for corroborating her own judgment of each student and for ranking students in the class. She stated that she would continue to use it for these purposes.

IV. Published Reviews

Miriam M. Bryan, Associate Director of Test Development, Educational Testing Service, Princton, N. J., devotes the first part of her review to a detailed and comprehensive history of the Stanford series. Although limitations of space prevent summary of this study here, we strongly recommend that it be consulted by anyone considering adoption of the Series.

Miss Bryan describes the 1964 Revision as "the product of five years of research and developmental work" (p. 115). It differs from the 1953 Revision in the following respects: Organization into five rather than four batteries provide[s] better at-grade coverage of content and skills" (p. 116). All items, with the exception of a few very simple ones at the Primary I level, are new. The same battery now includes a word reading test "which measures ability to analyze a word without the aid of context clues" (p. 116). A Word Study Skills Test has been introduced at the Primary I, Primary II, and Intermediate I levels; a separate Word Study Skills Test is now included with the Language Test. Miss Bryan describes the language tests as "carefully prepared, meticulously presented" to both pupil and teacher, and "organized logically and completely" (p. 116). She wonders, however, if this thoroughness has gotten out of hand and raises three questions concerning it. First, "is it necessary . . . to fragment the language testing process into so many parts above the Primary I level?" (p. 116). In Miss Bryan's opinion, such fragmentation "supports a

theory of language learning that may be less than logically or psychologically sound, and that must make language instruction the dull thing that it has become for great numbers of pupils" (p. 117). Secondly, are these power tests or speed tests? The reviewer suspects the latter and implies disapproval of such tests. Finally, "aside from the validity referred to in the manual as '*content*, or *curricular*, validity,' has any attempt been made to relate these separate measures to other evaluations of language skill?" (p. 117). As far as Miss Bryan is concerned, "*what* has been acquired of communication skill . . . , rather than *how* the acquisition has been accomplished" (p. 117), should be tested.

Miss Bryan's additional criticisms of the language tests are minor, but should be mentioned. She considers the absence of a listening test at the Intermediate I level a serious lack, and questions the correlation between a pupil's score on the spelling test and his spelling performance in free writing. She wonders whether usage should be treated "with the instruction to decide upon the basis of 'standard written English' " (p. 118). Other objections pertain to item type, such as those exercises which require students to select an answer from a list of given choices rather than to locate errors themselves; and the confusing instruction to draw a line through the correct response in the Paragraph Meaning Test at the Primary I and Primary II levels. However, Miss Bryan concludes her assessment of the language tests with the following statement: "In spite of all these questions and comments, the language tests remain impressive" (p. 118).

The remainder of Miss Bryan's comments pertain to reliability and validity, supplementary materials, and scoring, all of which she finds adequate. She regrets, however, that percentile ranks and stanines still appear to be geared to grade norms. She suggests that publication of additional information on the development of the series, the standardization sample, reliability and validity, intercorrelation among subtests, and item difficulty values "be placed high on the publisher's priority list" (p. 123). Despite her objections, she concludes by rating the *Stanford Achievement Tests* "high among standardized achievement test batteries designed for use at the elementary level" (p. 123).

[Note: Another review of the SAT by Robert E. Stake and J. Thomas Hastings appears in *The Sixth Mental Measurements Yearbook* and may be of interest to the reader. In general, these reviewers' conclusions coincide with those of Miss Bryan.]

Conclusions

The evidence of this study points overwhelmingly to the fact that there is no perfect objective test of English, nor does any currently published test come close to the goal of measuring success in English. The most that can be said for such tests is that a number of them measure rather well certain specific aspects of English, when these aspects are carefully selected as being valid elements in the English curriculum. The more concrete items of English, such as spelling, punctuation, capitalization, abbreviations, and other mechanical matters are tested with considerable success. English usage in specific words, phrases, and idioms, can be tested with reasonable success, provided that the items selected for testing are valid in terms of current English. Some tests in this survey have been criticized for the use of invalid usage items. Sentence structure ability, except of the simplest variety, is measured with difficulty and a great deal of unreliability. The infinite number of possibilities in the structure of sentences of ten or more words defies organization into any kind of objective testing. Finally, and of greatest significance to administrators and directors of educational testing, this survey reveals no evidence to support the hypothesis that composition, i.e., the art of writing in its entirety, can be measured by any objective test or any combination of such tests. It is, therefore, a sound conclusion of this survey that English as a school subject is not completely amenable to objective testing, but that some skills in the use of English may be tested. It follows that the school administrator and educational test director must know what is being tested in a so-called "English test" and interpret the result not as a measure of "English" but as the measurement of a limited number of English skills.

Before selecting an English test for general use in a particular school or school system, the administrator or test director should ask and find satisfactory answers to these questions:

1. What portions of the content of English at the grade levels to be tested are included in this test?
2. Is this proportionate emphasis parallel to the emphasis given by our teachers?

3. Does this test measure what our teachers consider to be a basic part of their instruction? In other words, does it truly test our curriculum?
4. Are the presented items valid? For example, are the items of usage, punctuation, sentence corrections, and other details consistent with what we teach?
5. What is the time required for this test?
6. How easy is it to administer? Are the directions simple and clear?
7. How easily may the test be scored?
8. What do the scores mean when completed?
9. How are the norms derived? How extensive was the sampling?
10. How can the results of this test be followed up for the improvement of the English program?

These questions may seem too minute or too many. It is true that they are not easy to answer without spending time on the study and analysis of tests. But in the light of the cost of testing, the time consumed in the administering and scoring of tests, and the psychological effects upon teachers and pupils in the giving and interpreting of tests, it is unsound pedagogy to administer tests in English until these questions can be satisfactorily answered.

Recommendations

From the foregoing analyses, some tests, in comparison with others, emerge from the critical process as being relatively more sound in content, better adapted to school use, more reliable in the interpretation of scores, and generally more satisfactory to the teachers who use them. The fact that a test is listed below does not mean that it is endorsed by the authors of this study. The potential user is referred to the analysis of the test to determine for himself its strengths and weaknesses, and the opinions of teachers who have used it. Similarly, omission of a test from the list below does not thereby suggest that the test should not be used. What is indicated is that the administrators and teachers of Wisconsin did not find it as useful and as satisfactory as the tests that are listed.

These six tests are suggested as useful in the schools of Wisconsin:

1. Cooperative English Tests (pp. 19-23)
2. School and College Ability Test (pp. 61-65)
3. Science Research Associates: Language Tests (pp. 65-69)
4. Sequential Tests of Education Progress: Essay Test (pp. 73-79)
5. Sequential Tests of Education Progress: Writing Test (pp. 73-79)
6. Stanford Achievement Test (pp. 79-85)

As a final word, this study urges that teachers and administrators give much thought to the problems of testing in English, in order to determine in advance exactly what is to be tested and to find the test which comes closest to meeting a particular need. By using such necessary precautions, testing in English can become a reasonably useful educational tool, rather than a haphazard casting of a net.

Test Publishers

The following is a list of the publishers and their addresses for each test used in this study:

Barrett-Ryan-Schrammel English Test, New Edition
Harcourt, Brace and World, Inc.
Test Department
757 Third Avenue
New York, New York 10017

California Language Test
California Test Bureau
Del Monte, Research Park
Monterrey, California 93940

Cooperative English Tests, 1960 Revision
Cooperative Test Division
Educational Testing Service
Princeton, New Jersey, or Los Angeles, California 90027

Differential Aptitude Tests
The Psychological Corporation
304 East 45th Street
New York, New York 10017

Essentials of English Tests, Revised Edition
American Guidance Service, Inc.
720 Washington Avenue S.E.
Minneapolis, Minnesota 55414

Greene-Stapp Language Abilities Test
Harcourt, Brace and World, Inc.
Test Department
757 Third Avenue
New York, New York 10017

Iowa Tests of Basic Skills
Houghton Mifflin Company
53 West 43rd Street
New York, New York 10036

Iowa Tests of Educational Development
Science Research Associates, Inc.
259 East Erie Street
Chicago, Illinois 60611

Metropolitan Achievement Tests
Harcourt, Brace and World, Inc.
Test Department
757 Third Avenue
New York, New York 10017

Objective Test in Grammar
The Perfection Form Company
214 West Eighth Street
Logan, Iowa 51546

Purdue High School English Test
Houghton Mifflin Company
53 West 43rd Street
New York, New York 10036

(Cooperative) School and College Ability Tests
Cooperative Test Division
Educational Testing Service
Princeton, New Jersey, or Berkeley, California 94704

Science Research Associates: Language Arts
Science Research Associates, Inc.
259 East Erie Street
Chicago, Illinois 60611

Science Research Associates: High School Placement Test
Science Research Associates, Inc.
259 East Erie Street
Chicago, Illinois 60611

Sequential Tests of Educational Progress
Cooperative Test Division
Educational Testing Service
Princeton, New Jersey, or Los Angeles, California 90027

Stanford Achievement Test
Harcourt, Brace and World, Inc.
Test Department
757 Third Avenue
New York, New York 10017

Acknowledgments

Special thanks are due to Dr. Oscar K. Buros for his generous permission to quote from *Mental Measurements Yearbook, Sixth Edition,* and earlier editions, and for his continuous interest in this project.

The Wisconsin English-Language-Arts Curriculum Project wishes to express its gratitude to the following persons for their invaluable assistance in preparing this study:

Elementary School:

Mr. Harold Billings, Ladysmith
Mrs. Elizabeth Bitzan, Kenosha
Mrs. Catherine Brunker, Amherst
Mrs. Hazel Chapman, Hudson
Mrs. Madeline Downer, Cadott
Mrs. Margaret Johnson, Stevens Point
Miss Genevieve Kalile, Boscobel
Miss Violet Littlefield, Sheboygan Falls
Mr. Archie C. Marten, Horicon
Miss Olga Martin, Eau Claire
Miss Monica McCabe, Oak Creek
Mrs. Marguerite Ryan, Prairie du Chien
Miss Ruth Saemann, Kohler
Mrs. Fern Spafford, Spooner
Dr. Reynold A. Swanson, Green Bay
Miss Helen Welling, Fond du Lac
Mrs. Doris White, Barron
Miss Helen Wrchota, Wisconsin Dells

Secondary School:

Mr. Robert Ademino, Spooner
Mrs. Margaret Arneson, Menomonie
Mr. Frederic B. Baxter, West Bend
Miss Marie Cahill, Eau Claire

Mrs. Lillie Carlson, Barron
Miss Margaret Gorr, Ashland
Mr. H. J. Knaus, Oshkosh
Mrs. Diane Lindenau, Cudahy
Mr. Mark A. Megna, Shawano
Mr. Dennis Nelson, Stevens Point
Miss Marilyn Radke, Milwaukee
Miss Irna Rideout, La Crosse
Mrs. Ruth F. Rosenthal, Menomonee Falls
Mrs. Herbert S. Roswell, Mauston
Mrs. Fern Stefonik, Rhinelander
Miss Marie Stepnoski, Fond du Lac
Miss Hazel Thomas, Milwaukee
Miss Emily Timmons Kenosha
Mr. Robert G. Vermillion, Oconomowoc
Miss Margaret E. Zielsdorf, Wausau
Mr. Victor Zimmerman, Ripon